Introduction to Onomap

Onomap is a research methodology, based on an academic project, which allows users to classify any list of names into groups of common cultural ethnic and linguistic origin using surnames and forenames.

There are growing needs to understand the nature and detailed composition of ethnic groups in today's increasingly multicultural societies. Ethnicity classifications are often hotly contested, but still greater problems arise from the quality and availability of classifications, with knock on consequences for our ability meaningfully to subdivide populations. Name analysis and classification has been proposed as one efficient method of achieving such subdivisions in the absence of ethnicity data, and may be especially pertinent to public health and demographic applications. However, previous approaches to name analysis have been designed to identify one or a small number of ethnic minorities, and not complete populations.

Onomap methodology is based on a new ontology of ethnicity that combines some of its multidimensional facets; language, religion, geographical region, and culture encapsulated in people's names. It is a methodology developed using data collected at very fine temporal and spatial scales, and made available, subject to safeguards, at the level of the individual. Such individuals are classified into 185 independently assigned categories of cultural ethnic and linguistic (Onomap) groups, termed Onomap Types, based on the probable origins of their names.

To read more about the Onomap methodology download a UCL Working Paper:

Mateos, Webber and Longley (2007) *The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods using Personal Names* , CASA Working Paper 116, Centre for Advanced Spatial Analysis, University College London

To read more about Name-based ethnicity classification methods, see a published review paper

Mateos, P. (2007) A review of name-based ethnicity classification methods and their potential in population studies, *Population Space and Place* , 13 (4): 243-263

Main Menu

The main menu of the Onomap software is comprised of the following options:

- Classify
- Statistics
- Unclassified
- Settings
- Exit Onomap

Classify

Load and code a names list with the Onomap classification.

This section constitutes the core of Onomap software and allows a user to automatically classify any list of surnames and forenames with a unique cultural ethnic and linguistic code termed Onomap Type, and a series of other attributes. It involves a process of seven steps.

For more details about this option see the names classification section.

Statistics

The statistics section offers a series of bar charts summarizing the frequency distribution of the last list of names classified by Onomap Group , by Religion , and by Geographical Area of origin, according to the specified Onomap Taxonomy .

This option is only available after a names list has been classified by Onomap.

Unclassified

The unclassified section produces a file with the surnames or forenames in the user file that Onomap was unable to classify with an Onomap Type. The user can save this file and send it to the Onomap team for further analysis of the reasons why they remained unclassified.

It is important to emphasise that only aggregated instances of surnames or forenames are separately reported, and full names of individuals never leave the user's computer.

This option is only available after a names list has been classified by Onomap.

Settings

In this screen the user can specify a customized working directory where all Onomap input and output files will be kept.

By default the working directory is placed in the location where Onomap software is running from.

Exit Onomap

Exits Onomap and closes the programme

Name classification process

The name classification process involves a series of screens where the user specifies the preferences for the automatic classification of names. This process consists of one initial screen called "Name Classification Process", and seven steps marked 1 to 7. At the end of the process the input list of names will be classified according to the user's preferences.

The initial screen explains this process and asks the user to provide two items of information:

- Which country is your name list from?

The user should specify the country from which the list of names have been gathered (this is of course independent from where the actual names are originated from). For example, if the list of names refer to a patient register in the UK, we will select United Kingdom, while if the list was collected from the US telephone directory, we will select United States.

- Apply name data cleansing and standarisation algorithm

The user should specify whether the name data cleansing and standarisation algorithm should be applied to their input file data before an attempt to classify the names is made.

Onomap name data cleansing and standarisation algorithm involves a series of rules and lookup tables of common misspellings, name variations and different naming conventions rules in order to maximise the chances of a match between the user's input data and Onomap name to ethnicity dictionaries. This can make a big difference in reducing the amount of names that remain unclassified.

As a result of the name data cleansing and standarisation algorithm, Onomap produces a 'standarised' version of a surname or forename that can be also appended to the export file for further analysis and enhancement.

Related links:

Standardised Surname and Forename

Step 1 of 7 - Provide a names list file
Step 2 of 7 - Visualise the input file records
Step 3 of 7 - Select the content of each column in the file
Step 4 of 7 - Select the columns to be exported
Step 5 of 7- Select additional Onomap fields to be exported
Step 6 of 7 - Specify export file name and location
Step 7 of 7 - Coding names and exporting file

Step 1 of 7 - Provide a names list file

Specify the location of the input file with the names of the people to be classified by Onomap will be supplied at this step. This input file will be in character separated values format (.csv) and should have been previously prepared following the instructions offered in the input file description page.

Step 1 of 7: File formatting problems found

Onomap performs a series of data consistency checks on the input file before it proceeds to its classification. Two possible errors in the data are detected at this stage:

- Inconsistent number of commas/columns

Onomap assumes that the first row of the input file contains the field names of the data. If the number of fields in the first row (specified by the number of commas in the csv file) does not match the number of fields in the rest of the file, an error message will appear. This is most likely to be caused by commas that appear within the input text, which can be prevented by specifying that the "text is enclosed by double quotes" when creating the csv file.

At this stage the user can decide whether to:

- View the problematic records to quickly visualise the root of the problem
- Save the problematic records in a file for further analysis outside Onomap
- Go Back to the previous screen, and specify a different input file
- Go to the Next screen, and decide to ignore this error message and carry on with attempting the classification leaving the input file as it is.

The user is highly encouraged to produce new csv files selecting the option to have text always enclosed by double quotes. Most software packages that save to csv format will include this option.

- Blank lines in the data

Onomap detects any number of rows that have no data or that are blank (no data between commas). The user can decide to remove these blank lines from the input file, or click Next, after which Onomap will ignore these blank lines.

However, there might be a different number of rows between the output and the input files as a consequence of ignoring blank lines. It is therefore highly encouraged that the user eliminates any blank lines in the input file before loading it in Onomap.

Step 2 of 7 - Visualise the input file records

This screen shows the contents of the first 50 records of the input file, for the user to familiarise with its contents.

The total number of records in the input file appears in the first line of the text.

Onomap assumes that the first line in the input file contains the column names.

The user can scroll up and down the screen to visualise the 50 records and also the column width can be changed to read longer text in columns.

Step 3 of 7 - Select the content of each column in the file

This screen helps Onomap to understand what is what in the information contained in the input file, that is, matching the user's columns to the data fields that Onomap can understand. Onomap will try first to match them automatically by using a 'best guess' approach to the users' column names.

- The appropriate columns that should be imported from the input file should be ticked

- The corresponding column name from the input file should be selected alongside each field name so that Onomap can match them.

- Only the fields Unique ID and Surname are mandatory, but it is highly recommended that you also include the Forename field if it is available, since this will result in a much more robust Onomap classification.

Step 4 of 7 - Select the columns to be exported

From this screen onwards the user will specify the configuration of the resulting export file with the names classified by Onomap.

Through this screen the user is allowed to include as many columns from the input file as she or he wishes.

- The column names on the left block are all the columns available in the user's input file

- The column names on the right block are those columns from the user's input file that will be exported into the export file, alongside further data that will be specified in the next step.

- The user will click on the arrows between the two blocks to select (> or >>) or deselect (< or <<) the columns as required.

- The only mandatory column in this step is that which represents the Unique ID, so that the user can then link the export file back to her/his original data.

Step 5 of 7- Select additional Onomap fields to be exported

This screen allows the user to add a series of Onomap fields to the export file

The user can select from a list of 16 additional Onomap classification fields:

- Onomap Type Code
- Onomap Type
- Onomap Subgroup
- Onomap Group
- People GB
- Geographical Area
- Religion
- Major Language
- Major Language ISO Code
- Major Language Family Tree
- 1991 Census Ethnic Group
- 2001 Census Ethnic Group
- 2001 Census Religion
- 2001 Census COB
- Onomap Score
- Onomap Coding Case
- Standardised Surname and Forename

- Only the Onomap Type Code and Onomap Type are mandatory fields in the export file, since they contain the core of the Onomap classification of the user's names.

Click on the above field names to read the definition of each of them.

Read more about the Onomap Taxonomy

Read more about the Onomap Classification Algorithm

Step 6 of 7 - Specify export file name and location

Specify the location of the export file which will contain the information specified in the previous two screens (4 and 5). The export file will be saved at the end of the process in character separated values format (.csv).

Onomap has now everything it needs to code the names by the Onomap classification and export the results to the export file. Click Next to initiate the coding.

Step 7 of 7 - Coding names and exporting file

This screen shows the status of the name-to-Onomap coding and file exporting process.

- The screen shows the number of records being coded every 0.5 seconds and the progress bar specifies the percentage of records classified.

- The coding process can be cancelled by clicking the “Cancel” button. If the process is cancelled the user can go back to change the options of the previous screens or can abort the process altogether and return to the Main Menu.

- When the coding process is finished, the rest of the buttons became activated and the user can choose to return to the Main Menu or go directly to Statistics section.

- The export file has now been saved in the location specified by the user

Statistics

The statistics section offers a series of bar charts summarizing the frequency distribution of the last list of names classified by Onomap Group , by Religion , and by Geographical Area of origin, according to the Onomap Taxonomy.

The statistics section is only available after a names list has been classified by Onomap.

Select Base

• The statistics for the users' names can be compared with a base for the distribution of names by Onomap groupings in the overall population at various geographical levels:

- Country

- Region

- Local

Read more on how the base is calculated.

The user can select only one base at any of these geographical levels, or decide not to compare their data by leaving the default base of "none". These three options represent different summaries of the results of the Onomap classified user names list, grouped by different criteria according to the Onomap Taxonomy. To read more about what each of these criteria mean, click on each of them in the paragraph above.

Select Groupings

- There are three different types of statistical summaries provided through a frequency histogram, that can be selected through a menu on the left of the screen:

• Statistics by Onomap Group

• Statistics by Religion

• Statistics by Geographical Area

These three options represent different summaries of the results of the Onomap classified user names list, grouped by different criteria according to the Onomap Taxonomy. To read more about what each of these criteria mean, click on each of them in the paragraph above.

Saving the Statistics

• The actual statistics section figures that lie behind the histograms displayed can be saved in a file with comma separated values (.csv). Click the "Save" button to specify the file location.

Saving the Statistics

• The actual statistics section figures that lie behind the histograms displayed can be saved in a file with comma separated values (.csv). Click the "Save" button to specify the file location.

Related Links:

• Frequency histogram

• Base for statistical comparisons

Frequency histogram

The main body of the statistics section depicts the results of the Onomap classified names list through a chart with a frequency histogram.

- The initial histogram shows total number of records in the input file, while when a base for statistical comparison is chosen, the histogram shows relative percentages.

- The full name of each bar and the percentage of records can be visualised by placing the cursor on over each bar (no need to click)

Base for statistical comparison

The base drop down menus allow to compare the user's name distribution with the average distribution against a general population base for a selected geographical level of dissagregation.

• This base has been calculated by classifying with Onomap the whole electoral register or telephone directory of each country. The electoral register contains all adults entitled to vote and is deemed to be the best names register available in most countries, while the telephone directory includes all "heads of households" that have registered their telephone line in the public directory.

• The user can select to compare her or his data against a country's national average, or a specific Region or Local area, according to official administrative groupings of local areas.

Onomap Settings Menu

In this screen the user can specify a customized working directory where all Onomap **input** and **output** files will be kept.

By default the working directory is placed in the location where Onomap software is running from.

Review Unclassified Names

The unclassified section produces a file with the surnames or forenames in the user input file that Onomap was unable to classify with an Onomap Type.

The unclassified section is only available after a names list has been classified by Onomap.

• The user can save this file and send it to the Onomap team for further analysis of the reasons why they remained unclassified.

There are two options to send the unclassified names file to the Onomap team as feedback:

- Via e-mail: will open up your default e-mail application and ask you to manually locate and attach the unclassified names file

- Via internet: will open up an html page resident in your computer, and then ask you to click a button to submit the unclassified names over the internet.

• It is important to emphasise that only aggregated instances of surnames or forenames are separately reported, and full names of individuals never leave the user's computer.

Input File

The input file with the names of the persons to be classified by Onomap needs to be provided with some minimum contents and in a particular data format. Please read these instructions carefully before attempting to classify your first file.

*** Input file contents**

- The input file will include one record per person to be classified, with a minimum of two fields for each record: Unique Identifier and surname.

- A Unique Identifier should be provided that unequivocally identifies each record in the file. This can be a sequential number, or any other alphanumeric identifier that exists only once in the file, hence a unique identifier of each record. This identifier will allow the user to link the resulting Onomap classified names back to the original dataset.

- At least the surname of each person should be provided in the input file. If the forename is also provided the resulting Onomap classification of the names will be much more robust.

- The file can contain any other data that the user wants to keep together for further analysis.

*** Input file format**

- The input file must be previously saved as a text file with extension '.csv' for *character separated values* format

- The columns in the input file must be separated by commas.

- The first line in the file should contain the column names for that file

- It is recommended that the columns containing text are enclosed by double quotes to make sure that if commas are present within the text a column is not split into separate fields.

- These three options (commas, column names, double quotes) are available in most software packages when exporting a file into '.csv' or *character separated values* format.

For example:

The input file in '.csv' format will look like:

"Unique Identifier", "Forename", "Surname"

6879037486, "Pablo", "Mateos"

6879037487, "John, Paul", "Smith"

6879037488, "Alex", "Singleton"

And Onomap will read it as :

Unique Identifier	Forename	Surname
6879037486	Pablo	Mateos
6879037487	John, Paul	Smith

6879037488	Alex	Singleton

Related links:

Unique Identifier

Surname

Forename

Export File

The main Onomap export file contains the result of classifying the names on the input file with the Onomap classification.

*** Export file contents**

The contents of the export file are defined by the user through the step 4 and step 5 of the coding process.

These can include some or all of the following fields - (*) denotes mandatory fields - :

- (*) Unique ID
- Surname
- Forename
- Other User input columns
- (*) Onomap Type Code
- (*) Onomap Type
- Onomap Subgroup
- Onomap Group
- People GB
- Geographical Area
- Religion
- Major Language
- Major Language ISO Code
- Major Language Family Tree
- 1991 Census Ethnic Group
- 2001 Census Ethnic Group
- 2001 Census Religion
- 2001 Census COB
- Onomap Score
- Onomap Coding Case
- Standardised Surname and Forename

*** Export file format**

The export file is comprised of a text file with extension `.csv' for *character separated values* format with the following characteristics:

- The columns in the input file will be separated by commas.

- The first line in the export file will contain the field names for the exported records.

- All fields containing text are enclosed by double quotes (e.g. “text”)

- These three options (commas, column names, double quotes) can be specified when reading the `.csv' export file into most software packages, to ensure a smooth transfer of data.

Related links:

Step 4 of 7 - Select the columns to be exported

Step 5 of 7 - Select additional Onomap fields to be exported

Unique ID

Unique ID stands for Unique Identifier.

In Onomap, a Unique ID unequivocally identifies each record in a names dataset. This is necessary since there could be more than one record with the exact same name in the input file (due to several people having the same name or several records of the same person).

- The Unique ID is provided by the user through the input file.

- The Unique ID can be any sequential number (like a record counter), or any other alphanumeric identifier that exists only once in the file (e.g. BROWNGORDON/09/2007)

- The Unique ID is always included in the export file and will allow the user to link the resulting Onomap classified names back to the original dataset in a convenient and accurate manner.

Surname

A surname, also known as family name or last name, normally represents the name shared by all the members of a family and usually corresponds to the component of a person's name inherited from his or her family.

Therefore, this is the key piece of information that allows Onomap to classify the probable cultural ethnicity and linguistic origins of a person.

Forename

A forename, also known as first name, given name, or Christian name, refers to the proper name given to a person usually at birth.

In Onomap, if a person has other forenames other than the main one, which are then known as middle names, these are never included in the forename component of a name.

Onomap Taxonomy

The Onomap classification is a methodology that allows users to classify any list of names into groups of common cultural ethnic and linguistic origin using surnames and forenames.

The classification of names into such groups of common origin is organised in a hierarchy of three levels of cultural ethnic and linguistic groupings, termed the Onomap Taxonomy.

At the base of such hierarchy there are a total of 185 independently assigned categories termed Onomap Types, which represent the smallest building blocks of the Onomap Taxonomy. These are then organised in 66 Onomap Subgroups which themselves nest together in much coarser 16 Onomap Groups.

In addition to these three levels, each Onomap Type has a series of attributes describing it that can also act as a way of aggregating these building bricks according to the different dimensions of a person's identity. These attributes are also part of the Onomap Taxonomy and include the following fields:

Geographical Area

Religion

Major Language

Major Language ISO Code

Major Language Family Tree

1991 Census Ethnic Group

2001 Census Ethnic Group

2001 Census Religion

2001 Census Country of Birth (COB)

You can group the Onomap Types according to one of these eight additional dimensions to analyse your results following different aspects of identity. Furthermore, you can even create your own aggregations of the 185 Onomap Types following a criteria that is more relevant to your analysis.

To facilitate these further aggregations you can view the whole Onomap Taxonomy click here to open a full table in text format.

Related links:

- Read a paper on alternative ethnicity groupings: *The Modifiable Ethnic Unit Problem*
- Export file
- Onomap Taxonomy Lookup Table

Onomap Type Code	Onomap Group	Onomap Subgroup	Onomap Type	Population GB 2004	Geograp
AF110	AFRICAN	AFRICAN	AFRICAN	4891	AFRICA
AF212	AFRICAN	AFRICAN	BOTSWANAN	8	AFRICA
AF214	AFRICAN	AFRICAN	MALAGASY	2	AFRICA
AF215	AFRICAN	AFRICAN	MALAWIAN	23	AFRICA
AF216	AFRICAN	AFRICAN	NAMIBIAN	1	AFRICA
AF218	AFRICAN	AFRICAN	SWAZILANDER	5	AFRICA
AF219	AFRICAN	AFRICAN	ZAIREAN	41	AFRICA
AF220	AFRICAN	AFRICAN	ZAMBIAN	274	AFRICA
AF221	AFRICAN	AFRICAN	ZIMBABWEAN	991	AFRICA
AF322	AFRICAN	AFRICAN	BURUNDIAN	18	AFRICA
AF324	AFRICAN	AFRICAN	KENYAN	1197	AFRICA
AF325	AFRICAN	AFRICAN	RWANDAN	18	AFRICA
AF327	AFRICAN	AFRICAN	TANZANIAN	104	AFRICA
AF429	AFRICAN	AFRICAN	BENINESE	7	AFRICA
AF430	AFRICAN	AFRICAN	CAMEROONESE	72	AFRICA
AF431	AFRICAN	AFRICAN	GAMBIAN	11	AFRICA
AF433	AFRICAN	AFRICAN	GUINEAN	15	AFRICA
AF434	AFRICAN	AFRICAN	IVORIAN	122	AFRICA
AF435	AFRICAN	AFRICAN	LIBERIAN	5	AFRICA
AF437	AFRICAN	AFRICAN	SENEGALESE	37	AFRICA
AF211	AFRICAN	BLACK SOUTHERN AFRICAN	BLACK SOUTHERN AFRICAN	5198	AFRICA
AF213	AFRICAN	CONGOLESE	CONGOLESE	1164	AFRICA
AF323	AFRICAN	ETHIOPIAN	ETHIOPIAN	1238	AFRICA
AF432	AFRICAN	GHANAIAN	GHANAIAN	46095	AFRICA
AF436	AFRICAN	NIGERIAN	NIGERIAN	88243	AFRICA
AF438	AFRICAN	SIERRA LEONIAN	SIERRA LEONIAN	6155	AFRICA
AF328	AFRICAN	UGANDAN	UGANDAN	1018	AFRICA
CL110	CELTIC	CELTIC	CELTIC	45653	BRITISH
CL211	CELTIC	IRISH	IRISH	3172876	BRITISH
CL212	CELTIC	IRISH	NORTHERN IRISH	223988	BRITISH
CL213	CELTIC	SCOTTISH	SCOTTISH	4749864	BRITISH
CL314	CELTIC	WELSH	WELSH	3065041	BRITISH
EA212	EAST ASIAN & PACIFIC	CHINESE	CHINESE	21185	EAST AS
EA218	EAST ASIAN & PACIFIC	CHINESE	MALAYSIAN CHINESE	3238	EAST AS
EA110	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	627	EAST AS
EA211	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	SOUTH EAST ASIAN	371	EAST AS
EA213	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	EAST ASIAN CARIBBEAN	2	AMERIC
EA221	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	SINGAPOREAN	583	EAST AS
EA225	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	TIBETIAN	13	EAST AS
EA316	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	INDONESIAN	116	EAST AS
EA319	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	BURMESE	1601	EAST AS
EA324	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	THAI	407	EAST AS
EA327	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	CAMBODIAN	59	EAST AS
EA328	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	LAOTIAN	120	EAST AS
EA414	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	FIJIAN	10	EAST AS
EA420	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	POLYNESIAN	54	EAST AS
EA422	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	SOLOMON ISLANDER	8	EAST AS
EA429	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	HAWAIIAN	2	EAST AS
EA430	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	MAORI	1	EAST AS
EA431	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	MAURITIAN	2	EAST AS
EA432	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	SAMOAN	10	EAST AS
EA433	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	TONGAN	9	EAST AS
EA434	EAST ASIAN & PACIFIC	EAST ASIAN & PACIFIC	TUVALUAN	2	EAST AS
EA215	EAST ASIAN & PACIFIC	HONG KONGESE	HONG KONGESE	119566	EAST AS
EA317	EAST ASIAN & PACIFIC	MALAYSIAN	MALAYSIAN	2092	EAST AS
EA323	EAST ASIAN & PACIFIC	SOUTH KOREAN	SOUTH KOREAN	2315	EAST AS
EA326	EAST ASIAN & PACIFIC	VIETNAMESE	VIETNAMESE	15723	EAST AS
EN315	ENGLISH	BLACK CARIBBEAN	BLACK CARIBBEAN	23665	AMERIC
EN110	ENGLISH	ENGLISH	ENGLISH	31118965	BRITISH
EN211	ENGLISH	ENGLISH	CORNISH	107068	BRITISH
EN213	ENGLISH	ENGLISH	CHANNEL ISLANDER	23995	BRITISH
EN314	ENGLISH	ENGLISH	BRITISH SOUTH AFRICAN	45	AFRICA
EU215	EUROPEAN	AFRIKAANS	AFRIKAANS	7805	AFRICA
EU734	EUROPEAN	ALBANIAN	ALBANIAN	3440	EASTER
EU727	EUROPEAN	BALKAN	BALKAN	16274	EASTER
EU729	EUROPEAN	BALKAN	BOSNIAN AND HERZEGOVIAN	1034	EASTER
EU730	EUROPEAN	BALKAN	MONTENEGRIN	44	EASTER
EU731	EUROPEAN	BALKAN	MACEDONIAN	371	EASTER
EU732	EUROPEAN	BALKAN	SLOVENIAN	1282	EASTER
EU733	EUROPEAN	BALKAN	CROATIAN	1362	EASTER
EU839	EUROPEAN	BALKAN	BULGARIAN	109	EASTER
EU620	EUROPEAN	BALTIC	BALTIC	0	EASTER
EU624	EUROPEAN	BALTIC	ESTONIAN	778	EASTER
EU625	EUROPEAN	BALTIC	LATVIAN	1559	EASTER
EU626	EUROPEAN	BALTIC	LITHUANIAN	1790	EASTER
EU836	EUROPEAN	CZECH	CZECH	4357	EASTER

EU837	EUROPEAN	CZECH	SLOVAKIAN	524	EASTER
EU212	EUROPEAN	DUTCH	BELGIAN FLEMISH	4417	CENTRA
EU214	EUROPEAN	DUTCH	DUTCH	20495	CENTRA
EU523	EUROPEAN	ENGLISH	MALTESE	8027	SOUTHE
EU110	EUROPEAN	EUROPEAN	EUROPEAN	31341	CENTRA
EU211	EUROPEAN	FRENCH	BELGIAN	815	CENTRA
EU213	EUROPEAN	FRENCH	BELGIAN WALLON	618	CENTRA
EU316	EUROPEAN	FRENCH	FRENCH	125754	CENTRA
EU317	EUROPEAN	FRENCH	BRETON	640	CENTRA
EU318	EUROPEAN	FRENCH	CANADIAN	299	AMERIC
EU319	EUROPEAN	FRENCH	FRENCH CARIBBEAN	3	AMERIC
EU420	EUROPEAN	GERMAN	GERMAN	129190	CENTRA
EU421	EUROPEAN	GERMAN	SWISS	128	CENTRA
EU838	EUROPEAN	HUNGARIAN	HUNGARIAN	11768	EASTER
EU522	EUROPEAN	ITALIAN	ITALIAN	229931	SOUTHE
EU835	EUROPEAN	POLISH	POLISH	155743	EASTER
EU840	EUROPEAN	ROMANIAN	ROMANIAN	744	EASTER
EU841	EUROPEAN	ROMANIAN	ROMANIAN BANAT	29	EASTER
EU842	EUROPEAN	ROMANIAN	ROMANIAN DOBREGA	28	EASTER
EU843	EUROPEAN	ROMANIAN	ROMANIAN MANAMURESCRIANA	331	EASTER
EU844	EUROPEAN	ROMANIAN	ROMANIAN MOLDOVA	200	EASTER
EU845	EUROPEAN	ROMANIAN	ROMANIAN MUNTENIA	364	EASTER
EU846	EUROPEAN	ROMANIAN	ROMANIAN TRANSILVANIA	835	EASTER
EU947	EUROPEAN	RUSSIAN	RUSSIAN	11118	EASTER
EU948	EUROPEAN	RUSSIAN	BELARUSIAN	27	EASTER
EU950	EUROPEAN	RUSSIAN	AZERBAIJANI	12	CENTRA
EU951	EUROPEAN	RUSSIAN	GEORGIAN	185	CENTRA
EU728	EUROPEAN	SERBIAN	SERBIAN	5279	EASTER
EU949	EUROPEAN	UKRANIAN	UKRANIAN	3948	EASTER
GR110	GREEK	GREEK	GREEK	30066	SOUTHE
GR211	GREEK	GREEK	GREEK CYPRIOT	79304	SOUTHE
HI110	HISPANIC	HISPANIC	HISPANIC	6084	SOUTHE
HI211	HISPANIC	PORTUGUESE	PORTUGUESE	86930	SOUTHE
HI212	HISPANIC	PORTUGUESE	BRAZILIAN	1949	AMERIC
HI213	HISPANIC	PORTUGUESE	ANGOLAN	458	AFRICA
HI214	HISPANIC	PORTUGUESE	GOAN	990	SOUTH A
HI315	HISPANIC	SPANISH	SPANISH	80180	SOUTHE
HI316	HISPANIC	SPANISH	CASTILLIAN	10775	SOUTHE
HI317	HISPANIC	SPANISH	LATIN AMERICAN	3644	AMERIC
HI318	HISPANIC	SPANISH	FILIPINO	1976	EAST AS
HI419	HISPANIC	SPANISH	BASQUE	1568	SOUTHE
HI520	HISPANIC	SPANISH	CATALAN	3105	SOUTHE
HI621	HISPANIC	SPANISH	GALICIAN	511	SOUTHE
IN110	INTERNATIONAL	INTERNATIONAL	INTERNATIONAL	15799	UNCLAS
JP110	JAPANESE	JAPANESE	JAPANESE	6335	EAST AS
JA313	JEWISH AND ARMENIAN	ARMENIAN	ARMENIAN	4353	CENTRA
JA110	JEWISH AND ARMENIAN	JEWISH AND ARMENIAN	JEWISH AND ARMENIAN	72	DIASPO
JA211	JEWISH AND ARMENIAN	JEWISH	JEWISH	80522	DIASPO
JA212	JEWISH AND ARMENIAN	JEWISH	SEPHARDIC JEWISH	821	DIASPO
ML427	MUSLIM	BANGLADESHI	BANGLADESHI	179401	SOUTH A
ML640	MUSLIM	ERITREAN	ERITREAN	1397	AFRICA
ML212	MUSLIM	IRANIAN	IRANIAN	10312	MIDDLE
ML216	MUSLIM	LEBANESE	LEBANESE	3107	MIDDLE
ML110	MUSLIM	MUSLIM	MUSLIM	103514	MIDDLE
ML428	MUSLIM	MUSLIM	MUSLIM INDIAN	25704	SOUTH A
ML431	MUSLIM	MUSLIM	MALAYSIAN MUSLIM	220	EAST AS
ML637	MUSLIM	MUSLIM	WEST AFRICAN MUSLIM	2399	AFRICA
ML639	MUSLIM	MUSLIM	SUDANESE	468	AFRICA
ML743	MUSLIM	MUSLIM	BALKAN MUSLIM	10	EASTER
ML211	MUSLIM	MUSLIM MIDDLE EAST	MUSLIM MIDDLE EAST	672	MIDDLE
ML213	MUSLIM	MUSLIM MIDDLE EAST	IRAQI	262	MIDDLE
ML214	MUSLIM	MUSLIM MIDDLE EAST	JORDANIAN	55	MIDDLE
ML215	MUSLIM	MUSLIM MIDDLE EAST	KUWAITI	3	MIDDLE
ML217	MUSLIM	MUSLIM MIDDLE EAST	OMANI	5	MIDDLE
ML218	MUSLIM	MUSLIM MIDDLE EAST	SAUDI ARABIAN	186	MIDDLE
ML219	MUSLIM	MUSLIM MIDDLE EAST	SYRIAN	142	MIDDLE
ML220	MUSLIM	MUSLIM MIDDLE EAST	UNITED ARAB EMIRATES	14	MIDDLE
ML221	MUSLIM	MUSLIM MIDDLE EAST	YEMENI	6	MIDDLE
ML530	MUSLIM	MUSLIM NORTH AFRICAN	MUSLIM NORTH AFRICAN	0	AFRICA
ML532	MUSLIM	MUSLIM NORTH AFRICAN	ALGERIAN	2585	AFRICA
ML533	MUSLIM	MUSLIM NORTH AFRICAN	EGYPTIAN	479	AFRICA
ML534	MUSLIM	MUSLIM NORTH AFRICAN	TUNISIAN	39	AFRICA
ML535	MUSLIM	MUSLIM NORTH AFRICAN	LIBYAN	38	AFRICA
ML536	MUSLIM	MUSLIM NORTH AFRICAN	MOROCCAN	572	AFRICA
ML320	MUSLIM	MUSLIM STANS	MUSLIM STANS	0	CENTRA
ML322	MUSLIM	MUSLIM STANS	KAZAKHSTANI	11	CENTRA
ML323	MUSLIM	MUSLIM STANS	KYRGYZSTANI	2	CENTRA

ML324	MUSLIM	MUSLIM STANS	TURKMENISTANI	8	CENTRA
ML325	MUSLIM	MUSLIM STANS	UZBEKISTANI	3	CENTRA
ML326	MUSLIM	MUSLIM STANS	AFGHANISTANI	3687	CENTRA
ML429	MUSLIM	PAKISTANI	PAKISTANI	508699	SOUTH A
ML430	MUSLIM	PAKISTANI KASHMIR	PAKISTANI KASHMIR	91472	SOUTH A
ML638	MUSLIM	SOMALIAN	SOMALIAN	33260	AFRICA
ML741	MUSLIM	TURKISH	TURKISH	50706	MIDDLE
ML742	MUSLIM	TURKISH	TURKISH CYPRIOT	1205	MIDDLE
ND211	NORDIC	DANISH	DANISH	20561	NORTHE
ND315	NORDIC	FINNISH	FINNISH	5685	NORTHE
ND110	NORDIC	NORDIC	NORDIC	6377	NORTHE
ND212	NORDIC	NORDIC	ICELAND	115	NORTHE
ND214	NORDIC	NORWEGIAN	NORWEGIAN	186375	NORTHE
ND213	NORDIC	SWEDISH	SWEDISH	19090	NORTHE
SK110	SIKH	SIKH	SIKH	283657	SOUTH A
SA214	SOUTH ASIAN	HINDI NOT INDIAN	HINDI NOT INDIAN	22106	SOUTH A
SA316	SOUTH ASIAN	HINDI NOT INDIAN	BANGLADESHI HINDI	2974	SOUTH A
SA211	SOUTH ASIAN	INDIAN HINDI	INDIAN HINDI	319677	SOUTH A
SA212	SOUTH ASIAN	INDIAN HINDI	INDIA NORTH	75282	SOUTH A
SA213	SOUTH ASIAN	INDIAN HINDI	INDIAN SOUTH	302	SOUTH A
SA110	SOUTH ASIAN	SOUTH ASIAN	SOUTH ASIAN	12699	SOUTH A
SA317	SOUTH ASIAN	SOUTH ASIAN	BHUTANESE	3	SOUTH A
SA318	SOUTH ASIAN	SOUTH ASIAN	NEPALESE	150	SOUTH A
SA420	SOUTH ASIAN	SOUTH ASIAN	SEYCHELLOIS	71	SOUTH A
SA421	SOUTH ASIAN	SOUTH ASIAN	KENYAN ASIAN	1121	AFRICA
SA522	SOUTH ASIAN	SOUTH ASIAN	ASIAN CARIBBEAN	581	AMERIC
SA523	SOUTH ASIAN	SOUTH ASIAN	GUYANESE	911	AMERIC
SA315	SOUTH ASIAN	SRI LANKAN	SRI LANKAN	53919	SOUTH A
ZU110	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	21826	NOT API
ZU111	UNCLASSIFIED	UNCLASSIFIED	NOT FOUND	110049	NOT API
ZZ110	VOID	VOID	VOID	12118	NOT API
ZZ211	VOID	VOID	VOID - SURNAME	819	NOT API
ZZ212	VOID	VOID	VOID INITIAL	94621	NOT API
ZZ213	VOID	VOID	VOID OTHER	56	NOT API
ZZ214	VOID	VOID	VOID - FORENAME	5464	NOT API
ZZ215	VOID	VOID	VOID TITLE	1858	NOT API

Onomap Type

An Onomap Type is a finely specified cultural ethnic and linguistic category created for the purpose of assigning a group of names to a common origin through a series of their shared characteristics.

- There are a total of 185 Onomap Types optimised for the contemporary population of Great Britain , and they conform the basic building blocks of the Onomap classification.

- Each Onomap Type has a unique Onomap Type Code

- Not all Onomap Types refer to the same level of granularity in the classification, some are very specific, such as 'EU212: BELGIAN FLEMISH' and others very general, such as 'ML110: MUSLIM', depending on how well a name's origin represents an Onomap Type.

- In practice, when a person's name is coded by Onomap each person is assigned with an Onomap Type, through the Onomap Classification Algorithm. The rest of the information appended to the export file relate to each individual Onomap Type following the relationships established in the Onomap Taxonomy.

- For a full list of the 185 Onomap Types open the Onomap Taxonomy Lookup Table

Related links:

Onomap Taxonomy

Onomap Taxonomy Lookup Table

Onomap Type Code

Onomap Classification Algorithm

Onomap Type Code

The Onomap Type Code uniquely identifies each Onomap Type.

It is comprised of five alphanumeric characters, an example being ‘AF211':

- The first two characters are two letters that identify the Onomap Group

- The second three characters are three numbers

The Onomap Type code is always included in the export file together with the Onomap Type name.

One advantage of using codes is that the user may choose for example to change the names of the Onomap Types with her or his own descriptions, while the Onomap Type Code in the classified datasets will always allow to create look-ups to the Onomap Taxonomy.

Onomap Subgroup

An Onomap Subgroup is an aggregation of one or more Onomap Types.

Onomap Subgrops have been produced by merging Onomap Types using the ` People GB' population counts until a minimum population size of 1,000 people could be met, so that all Onomap Subgroups would all have a minimum size. Furthermore, additional Onomap Types were merged when they were deemed very culturally similar for reporting purposes, for example; Galician, Castillian, Basque and Catalan Onomap Types were all merged into a Spanish Onomap Subgroup.

A total of 66 Onomap Subgroups were finally compiled, which all had sufficient internal consistency and a minimum size.

A full list of the Onomap Subgroups can be viewed in the Table of Onomap Groups and Onomap Subgroups

Onomap Group

Onomap Groups are aggregations of Onomap Types and Onomap Subgroups into 16 coarse groups of common origin usually at the subcontinental level. It represents the highest level of the hierarchy in the Onomap Taxonomy.

A full list of the Onomap Grops can be viewed in the Table of Onomap Groups and Onomap Subgroups

Table of Onomap Groups and Onomap Subgroups

Onomap Group	Onomap Subgroup
AFRICAN	AFRICAN
	BLACK SOUTHERN AFRICAN
	CONGOLESE
	ETHIOPIAN
	GHANAIAN
	NIGERIAN
	SIERRA LEONIAN
	UGANDAN
CELTIC	CELTIC
	IRISH
	SCOTTISH
	WELSH
EAST ASIAN & PACIFIC	CHINESE
	EAST ASIAN & PACIFIC
	HONG KONGESE
	MALAYSIAN
	SOUTH KOREAN
	VIETNAMESE
ENGLISH	BLACK CARIBBEAN
	ENGLISH
	AFRIKAANS
	ALBANIAN
	BALKAN
	BALTIC
	CZECH
	DUTCH
	ENGLISH
	EUROPEAN

Onomap Group	Onomap Subgroup
GREEK	GREEK
HISPANIC	HISPANIC
	PORTUGUESE
	SPANISH
INTERNATIONAL	INTERNATIONAL
JAPANESE	JAPANESE
JEWISH AND ARMENIAN	ARMENIAN
	JEWISH
	JEWISH AND ARMENIAN
MUSLIM	BANGLADESHI
	ERITREAN
	IRANIAN
	LEBANESE
	MUSLIM
	MUSLIM MIDDLE EAST
	MUSLIM NORTH AFRICAN
	MUSLIM STANS
	PAKISTANI
	PAKISTANI KASHMIR
	SOMALIAN
	TURKISH
NORDIC	DANISH
	FINNISH
	NORDIC
	NORWEGIAN
	SWEDISH
SIKH	SIKH

EUROPIAN	FRENCH
	GERMAN
	HUNGARIAN
	ITALIAN
	POLISH
	ROMANIAN
	RUSSIAN
	SERBIAN
	UKRANIAN

SOUTH ASIAN	HINDI NOT INDIAN
	INDIAN HINDI
	SOUTH ASIAN
	SRI LANKAN
UNCLASSIFIED	UNCLASSIFIED
VOID	VOID

People GB

People GB in the Onomap Taxonomy refers to the number of people in the Great Britain (GB) electoral register whose names have been classified into a particular Onomap Type.

This information is offered as an additional field to be included in the export file.

The electoral register contains all adults entitled to vote and is deemed to be the best names register available in Great Britain .

Geographical Area

Geographical Area in the Onomap Taxonomy refers to a broad subcontinental division of the world in which an Onomap Type falls.

There are 14 broad Geographical Areas, three of which are actually non-geographical

The level of disagregation is dictated by the number of distinct Onomap Types and their population weight in Great Britain.

GEOGRAPHICAL AREA
AFRICA
AMERICAS
CENTRAL ASIA
EAST ASIA
SOUTH ASIA
MIDDLE EAST
BRITISH ISLES
NORTHERN EUROPE
EASTERN EUROPE
SOUTHERN EUROPE
CENTRAL EUROPE
DIASPORIC
NOT APPLICABLE
UNCLASSIFIED

To see the full list of assignments between Onomap Types and Geographical Areas open the Onomap Taxonomy Lookup Table

Religion

Religion in the Onomap Taxonomy refers to the closest religious group identified for each Onomap Type.

There are some cases where the match is perfect, such is the Onomap Type ML110: Muslim and the Muslim religion, and other where the assignment is more speculative.

There are 12 religious groups in the Onomap Taxonomy to which an Onomap Type is assigned to.

Religion
BHUDDIST
CHRISTIAN
CHRISTIAN: CATHOLIC
CHRISTIAN: GREEK ORTHODOX
CHRISTIAN: ORTHODOX_CALCEDONIAN
CHRISTIAN: PROTESTANT
CHRISTIAN: RUSSIAN ORTHODOX
HINDU
JEWISH
MUSLIM
SIKH
NOT APPLICABLE

The religion to Onomap Type allocations where carried out checking each Onomap Type individually in different websites and dictionaries.

Two of the main on-line sources consulted are:

Country Profiles at Joshua project

CIA World Factbook

To see the full list of assignments between Onomap Types and Religions open the Onomap Taxonomy Lookup Table

The user can obviously decide to re-arrange the religion assignments of each Onomap Type to fit their own purpose.

Major Language

Major Language in the Onomap Taxonomy refers to the closest language identified for an Onomap Type. This usually refers to the majority language spoken within an Onomap Type .

In a majority of cases an Onomap Type is in itself equivalent to a language, such as JP110: Japanese, but in others the major language assignment is more speculative, selecting only the largest language for a population group or country in terms of number of speakers.

The user can obviously decide to re-arrange the language assignments to fit their own purpose.

The major language allocations where carried our checking each Onomap Type individually in different websites and dictionaries.

The main on-line sources consulted are:

Ethnologue

Country Profiles at Joshua project

CIA World Factbook

The final name given to a language have been taken from Ethnologue language classification, the most commonly accepted language classification of over 7,000 languages, all coded to standard linguistic codes.

- Read more about the Major Language ISO Code used by Onomap.

- See the full list of assignments between Onomap Types and Major Languages by opening the Onomap Taxonomy Lookup Table

Major Language ISO Code

The Major Language ISO Code is the international standardised code assigned to each Onomap Type's Major Language in the Onomap Taxonom. The ISO language code, has been introduced as part of the Ethnologue language classification, the most commonly accepted language classification of over 7,000 languages, all coded to standard linguistic codes. The Ethnologue coding system is now part of the International Standards Organisation as ISO 639, and this is the format adopted in Onomap.

To see the full list of Major Language ISO Codes used by Onomap open the Onomap Taxonomy Lookup Table

Related links:

Ethnologue

ISO 639

Major Language Family Tree

The Language Family Tree in the Onomap Taxonomy is the overall language family plus its main branches from which the Onomap Type's Major Language was originated, according to the Ethnologue classification (e.g. Indo-European).

Languages in Ethnologue classification can have a variable number of levels in their family tree (branches), in some cases up to ten levels. In the Onomap Taxonomy only the four coarser levels of the tree are included (closer to the 'trunk' of the tree). These four levels are included in the Onomap Taxonomy as concatenated text separated by semicolons. The language family tree can be read as branching out from left to right in the text.

For example, the language family tree for Punjabi is:

INDO-EUROPEAN;INDO-IRANIAN;INDO-ARYAN;NORTHWESTERN ZONE

Therefore, Punjabi belongs to a major language family named 'Indo-European', and within that to the 'Indo-Iranian' subfamily, 'Indo-Aryan' branch, and 'Northwestern Zone' subbranch.

By segmenting the tree at different levels, the user may choose to create their own classification of languages by aggregating the major language into different levels of language similarity, and hence aggregating the Onomap Types associated with them in the Onomap Taxonomy using different language grouping criteria.

To see the full list of Major Language Family Trees used by Onomap open the Onomap Taxonomy Lookup Table

1991 Census Ethnic Group

The 1991 Census Ethnic Group in the Onomap Taxonomy is the closest ethnic group in the UK 1991 Census Ethnicity Classification, which was comprised of the following 9 ethnic groups:

1991 UK Census Ethnic Group	
0	White
1	Black - Caribbean
2	Black - African
3	Black - Other
4	Indian
5	Pakistani
6	Bangladeshi
7	Chinese
8	Any other ethnic group
9	Not Given

The matching between Onomap Type and UK Census Ethnic Groups was done by using the criteria specified by the Office for National Statistics to assign write-in answers to the Census back into the official ethnic groups. These criteria are explained in:

Office for National Statistics. 2003. *Ethnic group statistics: A guide for the collection and classification of data.* Available at: http://www.statistics.gov.uk/about/ethnic_group_statistics/downloads/ethnic_group_statistics.pdf. Accessed: 13/02/2006.

2001 Census Ethnic Group

The 2001 Census Ethnic Group in the Onomap Taxonomy is the closest ethnic group in the UK 2001 Census Ethnicity Classification, which was comprised of the following 16 ethnic groups:

2001 UK Census Ethnic Group	
White	
A	British
B	Irish
C	Any other White
Mixed	
D	White and Black Caribbean
E	White and Black African
F	White and Asian
G	Any other mixed
Asian or Asian British	
H	Indian
J	Pakistani
K	Bangladeshi
L	Any other Asian background
Black or Black British	
M	Caribbean
N	African
P	Any other Black background
Other Ethnic Groups	
R	Chinese
S	Any other ethnic group
Z	Not stated

The matching between Onomap Type and UK Census Ethnic Groups was done by using the criteria specified by the Office for National Statistics to assign write-in answers to the Census back into the official ethnic groups. These criteria are explained in:

Office for National Statistics. 2003. *Ethnic group statistics: A guide for the collection and*

classification of data. Available at: http://www.statistics.gov.uk/about/ethnic_group_statistics/downloads/ethnic_group_statistics.pdf . Accessed: 13/02/2006.

2001 Census Religion

The 2001 Census Religion in the Onomap Taxonomy is the closest religion in the UK 2001 Census Religion Classification, which was comprised of the following 7 religions:

2001 UK Census Religion
BHUDDIST
CHRISTIAN
HINDU
JEWISH
MUSLIM
SIKH
OTHER

The matching between Onomap Type and UK 2001 Census Religion was done by assigning the Religion field in the Onomap Taxonomy to their corresponding major religious group in the UK Census.

2001 Census COB

The 2001 Census Country of Birth (COG) in the Onomap Taxonomy is the closest country or region of birth in the UK 2001 COB classification, which was comprised of the following 7 COB regions:

2001 Census COB
COB_ENGLAND
COB_NORTHERN IRELAND
COB_SCOTLAND
COB_REPUBLIC OF IRELAND
COB_OTHER EU COUNTRIES
COB_WALES
COB_ELSEWHERE

The matching between Onomap Type and UK 2001 COB was done by manually deciding which geographical region it was included.

Onomap Classification Algorithm

The Onomap Classification Algorithm is the set of rules by which Onomap software decides how an individual's name is assigned to one final Onomap Type and how the overall person's Onomap score is calculated.

Each person is due to have two components of a name, a forename and a surname. Both of these elements are evaluated against a 'Name-to-Onomap Type' dictionary that determines the likely Onomap Type of each of them.

When putting these two together, different cases are possible, such as that both elements of a person's name share the same Onomap Type, that they only share a common Onomap Subgroup or an Onomap Group, or that they do not share any of the three, in addition to other data handling situations.

In order to consider all of these posible cases when evaluating a person's Onomap Type, the Onomap Classification Algorithm is comprised of a set of seven rules or cases. At the end of the coding process, a person is assigned with an Onomap Type according to one of these seven cases defined as the Onomap Coding Case.

Read more about these seven rules in the Onomap Coding Case section.

Onomap Coding Case

The Onomap Coding Case is the actual rule or case in the Onomap Classification Algorithm with which a particular name was coded in Onomap.

There are eight rules that are defined as the Onomap coding cases, and that take care of all of the possible cases in the Onomap Classification Algorithm. The Onomap coding case with which a particular person has been allocated can be added as an output in the export file.

These eight rules or Onomap coding cases are evaluated for every individual person in a sequential order from Case 1 to Case 7. When a person's name meets the criteria specified in one case the Onomap Type is assigned and no further cases are evaluated.

CASE 1

Both the Surname AND Forename are unclassified or not found in the dictionaries

CASE2

The Surname OR Forename is unclassified or not found in the dictionaries

CASE 3

Both Onomap Types are the same: the person is assigned to that Onomap Type

CASE 4

Both Onomap Subgroups are the same: the person is assigned to that Onomap Subgroup

CASE 5

If the absolute difference between Onomap Scores of each name element is larger than 0.2: the person gets assigned to the Onomap Type with the highest score

CASE 6

If the Onomap Groups are the same: the person is assigned to that Onomap Group

CASE 7

If the absolute difference between Onomap Scores of each name element is smaller than 0.2: the person gets assigned to the Onomap Type with the highest score

CASE 8

If the forename and surname can not be identified because of formatting problems: the person gets assigned to the Onomap group: "Unclassified"

At the end of this process each person's full name will have an overall Onomap Type assigned to it, either at the level of Onomap Type, Onomap Subgroup or Onomap Group level, or remains unclassified.

Onomap Score

Each forename or surname has been assigned with an Onomap Type together with a score for that relationship, which represents the probability of a name having originated from a particular Onomap Type.

Onomap Scores vary between 0 (very low probability) and 1 (very high probability). At the end of the classification, a person is assigned with a final Onomap Type and an overall person's Onomap Score, which represents the probability of that person belonging to that Onomap Type.

Read the Onomap Classification Algorithm to read more about how the Onomap classification decides how an individual's name is assigned to which Onomap Type.

According to the Onomap Coding Case under which a person has been assigned to an Onomap Type the final person's Onomap score is calculated as follows:

a- CASE 3, CASE 4 and CASE 6 (coincident surname and forename Onomap Types, Subgroups or Groups): The final person score is the **average** of the surname and forename scores

b- CASE 5 and CASE 7 (divergent surname and forename Onomap Types, Subgroups or Groups): The final person score is the **absolute difference** of the surname and forename scores.

c- CASE 1 and CASE 2 (not found, unclassified or single name element): The final person score is **equal to 0**