

Getting Started Guide UK

Copyright

matchIT is copyright © helpIT systems inc. 1994-2011, all rights reserved.

FoxPro is copyright © Microsoft Corporation 1988-2011, all rights reserved.

Trademarks

matchIT is a registered trademark of helpIT systems inc.

FoxPro is a registered trademark of Microsoft Corporation.

All other trademarks are also acknowledged.

Table of Contents

Getting Started.....	1
This Guide	1
Installation	1
Evaluation System Limitations	1
matchIT v5 User Interface	2
The matchIT Toolbar	2
The File Selector Button	2
matchIT Wizard Pane	3
Exercise 1	4
Single File Deduplication	4
Importing Data.....	4
Single File Wizard Operation	4
Automatic Import Wizard.....	5
Finding Duplicates.....	7
Viewing Matches	8
Flagging Matches	10
File Output.....	11
Output File Layout.....	12
Conclusion	13
Exercise 2	14
Multiple File Deduplication.....	14
Starting the Two File Wizard	14
Find Suppressions.....	16
View/ Verify Overlap	17

Remove Suppressions	18
Generating a Clean List	18
QA Dashboard	19
Postcoding and Address Validation	22
Suppression/ Enhancement Service	24
Further Information	25

Getting Started

This Guide

matchIT consists of several optional modules, which are totally integrated into one package. Depending on the modules that you purchase (or are evaluating), options are enabled or disabled i.e. greyed out. The menu structure also depends on what modules are activated. This Getting Started Guide focuses mainly on use of matchIT for dedupe, merge/purge, salutations, casing and Mailsort . If any of these functions are not relevant to you, please ignore any instructions relating to the corresponding options. Similarly if you are using addressIT for postcoding or sortIT for Mailsort, simply select the relevant options when they appear during exercise 1. More information about addressIT is provided on page 22 and a separate guide is provided for sortIT.

Installation

To install matchIT, if installation does not start automatically, please run the SETUP program on the CD and follow the instructions given.

Note matchIT needs at least **1GB** of RAM (memory) in order to run reliably and **300MB** of free hard disk space for even small data files.

If you do not have an Activation USB stick, you will need to contact your supplier for an activation code before you can use matchIT to process anything other than the supplied example data. You can do this from the matchIT Evaluation screen as described in the next section.

If you are installing under Windows, you must have administrator rights to install matchIT and/or the matchIT activation code.

If some of the dialog boxes are not fully visible after you have installed matchIT, make sure your screen resolution is set to at least 1024 x768.

Evaluation System Limitations

The standard evaluation system is configured as matchIT Campaign with sortIT and is limited to 15 days and files of up to 5,000 records. Until activated, the matchIT Evaluation screen is displayed whenever you start matchIT.










Simply follow the instructions on screen if you want to activate matchIT at this stage, then Cancel and restart matchIT to use the activated version. If you do not wish to activate yet, just click on the Run button.

Note: Prior to activation, you will only be able to run matchIT on the example data supplied.


matchIT v5 User Interface

While all of matchIT's processes can be accessed from the menu bar, the most common selections have been integrated into v5's redesigned User Interface. Following is a description of each icon.

The matchIT Toolbar

	Open an existing matchIT File		View Text File
	Import Data		ODBC Connection
	Find Matches		Find Overlap
	Generate Output		matchIT Options
	Mailsort		





The File Selector Button

	There is frequent reference to this button in this guide.
---	---

matchIT Wizard Pane

The Wizard Pane gives the user immediate access to any of matchIT's file wizards. Included in the Pane is a quick link to the matchIT Job Script window. matchIT's file wizards have the ability to lead a user through the entire data cleansing process. The file wizards allow inexperienced users to work through data cleansing processes with ease, while still maintaining the advanced data cleansing functionality that has made matchIT an industry leader.

A description of each icon follows:

	Single File	Import and dedupe a single file.
	Two File	Work through various 2 file processes.
	Multiple File	Import, dedupe and merge two or more files simultaneously.
	Automation	Access matchIT's pre-programmed job scripts or create your own.

Exercise 1

Single File Deduplication

This section of the guide takes you through the process of deduplicating a single file. matchIT works on a copy of your data, which is in DBF format. To create this copy, you need to Import your data into matchIT.

Importing Data



The easiest and quickest way of importing data into matchIT is by using the Setup Wizard. You can start the wizard from the single file button in the Wizard Pane.

The wizard builds a database inside matchIT, into which your data is copied.

To begin with, you are asked which file you wish to import from. The Setup Wizard initially defaults to looking in the matchIT\Import directory, but a different drive or directory can be selected. The Setup Wizard helps to import any of the common text file types as well as native Access, Excel and ODBC data sources.

If you have a text file to import and don't see it listed, select All Files (*.*). Alternatively, if you select DBF Tables, Access Databases or Excel Worksheets from the Files of Type drop down list, the appropriate files are displayed. To use ODBC you must set up an ODBC connection first.

To let you follow through a worked example, we have included a file called EXAMPLE1.TXT in this directory. This is a file of test data, using fictional data made up for the purpose of demonstrating matchIT. This section of the guide assumes that you are using that file. You can, of course, use your own data instead; this guide will be most relevant to data that contains similar information to the example data, but you should be able to interpret these instructions as appropriate.

First, select the Single File button on the left hand side of the matchIT user interface. Then use the file selector button to open the file selector window.

Either highlight EXAMPLE1.TXT and click the Open button, or double click this file, then select Continue.

Single File Wizard Operation

First, the Setup Wizard will determine the type of data in each field. Here you have two choices:

Switch to Automatic Wizard

This is the default option. The wizard will automatically attempt to determine the data type. (See **Automatic Import Wizard**)

Switch to Manual Wizard

Choose this option if you want to manually specify what the various data items represent. This is usually appropriate if your data file is in a Fixed-Width format, unusual in its layout or content, or if the Setup Wizard has already failed to determine the data type. (See the **Manual Import Wizard** section in the **Going Further with matchIT** guide.)

Automatic Import Wizard

The Field Layout window is displayed, immediately following your selection of an import file. You should check that the Setup Wizard has correctly identified the contents of each name or address field, using the scroll bars to view more fields and records (the Continue button is not enabled initially, as the field names must be reviewed first).

If the Setup Wizard has not understood what is in a column correctly, right click on the field name (i.e. the column heading) and then select Rename Field. The appropriate field name can then be chosen from the drop down list. For all the name and address fields in your data, you should use the names that are shown in the field name drop-down list. matchIT refers to these field names as part of its processing, so all the address lines should simply be labeled ADDRESS LINE - matchIT will number them, and compare across all the address lines when checking for duplication.

If you are using your own customer data file and have provided matchIT with a unique reference number for each record in the file, you can right click on the column heading for this field and label it Unique_Ref - this will enable you to export reference numbers for matching pairs to a file. Enabling you to remove duplicate records from your source database. The data in the field must be genuinely unique within the input file, or else you will receive an error message after import.

If you have data items in your own data files that are not relevant to deduplication or mailing, you can enter your own field name by clicking on the Rename Field button. First, check that the top dropdown list (Change field labeled) shows the field name that you want to change. You can then type the appropriate field name in the *Enter New Field Name* box.

For Comma or Tab Delimited input files, the Setup Wizard uses a default width of 40 characters, which it increases if it thinks necessary. However, you should change this if the actual maximum width of that field is greater or significantly less than 40. In EXAMPLE1.TXT, you could change the width of TELEPHONE to 20 or 25 characters, or leave it at 40 characters if you are not worried about saving disk space. The Setup Wizard uses a sample size from the input file, which is the first 1,000 records (by default). To see the longest string of data in a field, you can scroll to the bottom of the record list to view a "record" highlighted in pink, which shows the longest example of each field in the file sample.

If the input file is a fixed width text file, you can use the Split Field or Combine Fields buttons to correct the Setup Wizard (if it has not been able to work out where one field ends and the next begins). When splitting fields, click in the data window at the point where the field should end, before you select Split Field. You will then have to use the Rename Field button to correct the field name.

matchIT Setup Wizard

File Information

File to import: [C:\... \MATCHITV51\IMPORT\EXAMPLE1.TXT](#)

File format: **Comma Delimited** Field count: **9**

This field should contain the FULL name, such as 'Mr John Smith'

ADDRESSEE	JOB_TITLE	COMPANY
MR G K CHILVERS	LASER MANAGER	STANTON COPSHAW DIREC
MR D G HOLMES MSC	PARTNER	BOWERS SYSTEMS
THE MANAGING DIRECTOR		SANFORD SERVICES
JOHN TAITLE	INFORMATION SYSTEMS MANAGER	FRINTON WILKES PLC
MR R J EAST	MARKETING MANAGER	MOTOR TRADERS SUPPLIE
MR CLIFTON DAVIES	OFFICE SERVICES CO-ORD	BRITISH PHILATELIC BU
MR LUKE GENTRY	SALES MANAGER	PLUMB BOOKING COMPANY
PETER WILLITS	MANAGING DIRECTOR	ADMIRAL SOUND SYSTEMS
SIR SIMON POLLOCK	CHAIRMAN	LUXURY TRAVEL CLUDES L
GILL LAUGHTON	COMMERCIAL MANAGER	GRETCH LTD
MR GREVILLE GIVENS		GIVENS ADVERTISING
MR RON BARTON LBIPP	MANAGING DIRECTOR	GIVENS ADVERTISING
MR V MACNICHOLS		WETHERALL CONSULTING
MAJOR J L BELKSTEIN		DAVIS MANAGEMENT SERV
CHRISTIAN LE MARCHANT	FINANCIAL MANAGER	D'ARCEY CONTROLS
PROFESSOR J JEFFERIES	SECRETARY / TREASURER	B.X.I. MARKETING LTD
DORIS MONK		PICNIC BASKET PLC
CHRISTINE BICKNELL	DIRECTOR	COWAN SUPPLIES
ANDY MCWILLIAMS	MARKETING CONSULTANT	ASKAN BUSINESS PUBLIC
MR W SUMMERS		
MR P DAWSON	FINANCIAL CONTROLLER	BENTLEY JENNISON
MR L P J TAYLOR		BRINTMORE INTERNATION

Field Information

Field name:

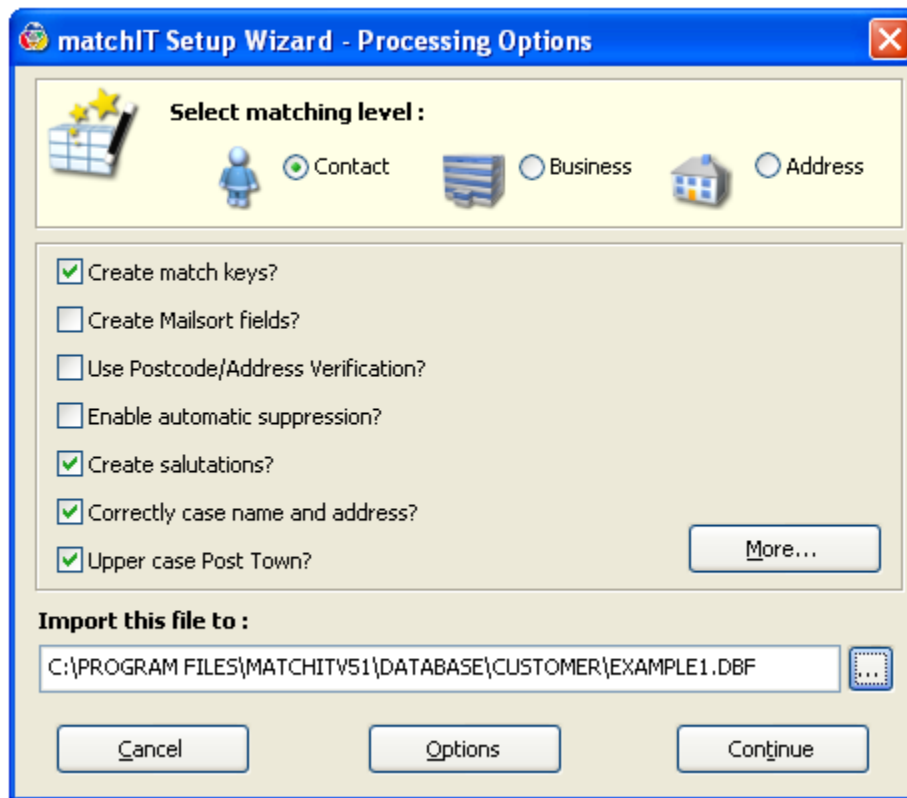
Field width:

You can use the *Add Field* button to create extra fields which *matchIT* can populate later on e.g. if your input name is of the form "Mr J Smith" or "John Smith" all in one field, you can add fields for Prefix, Forenames and Surname, which *matchIT* will populate automatically on Import with the appropriate components of the name.

When you have reviewed the field names given to your data, select *Continue*. The next dialog box displayed allows you to tell the Setup Wizard what to do with the data, after it has been imported.

Finding Duplicates

The next dialog box shown allows you to tell the Setup Wizard what needs to be done to the data after importation.



Select options as follows:

- **matchIT** needs to **Create match keys** if you are going to find duplicates, so this option will be selected by default, but you can deselect it if you do not want to dedupe this file or compare it with another file.
- If you wish to Mailsort the file then tick **Create Mailsort fields**.
- Click on Generate Additional Fields to open a window offering further features which matchIT can apply to your data during the import process. The options available cover creating salutations, generating quality score fields, generating PAF description fields (if you have addressIT) and options to split and combine name elements.
- **matchIT Pro** also allows you to "proper case" the name and address i.e. convert data intelligently to upper and lower case, so tick the **Correctly case name and address** option.
- You should decide on which level of deduplication your data needs, using the **Select matching level** section. Because this data has a COMPANY field, **matchIT** assumes this is business data, so the options are: dedupe down to one record per **Contact**, one record per **Business** or one record per **Address**. With consumer data the options are: dedupe down to one record per **Individual**, one record per **Family** or one record per **Household**. Leave the setting as 'Contact' for this exercise.

• **Options** allows you to change the many options that **matchIT** utilises, such as the default salutation, whether to automatically exclude records that have suspect data in them (e.g. if someone has entered "New address needed" in the address lines) and a host of lower level options.

• The file selector button will let you overwrite the default file name and destination. Enter a name of **WORKFILE**: this new file will be saved in DBF format.

For this example file, after the options have been set, select Continue. When you become more familiar with **matchIT**, you can specify different matching criteria from the default settings - see **Going Further with matchIT**.

Now *matchIT* will start processing. This involves:

- Importing the original data
- Generating the key fields for finding duplicates
- Enhancing the data as requested e.g. salutations, casing, relocating Postcode and country data to fixed fields
- Locating the duplicates within the file
- Generating first level reports.

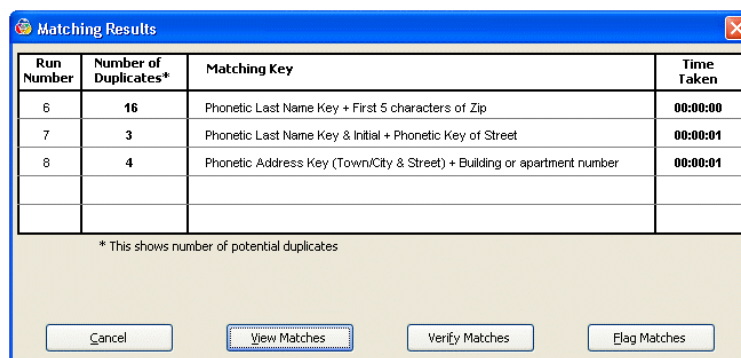
This stage should be very quick on the first example file, but is dependent upon file size, hardware, and software configurations. **matchIT**'s performance is benchmarked at a rate of several million records per hour.

Note There are many quick and simple things that you can do to speed things up. Some of them apply just to specific areas, such as finding matches - if you want to know more about tuning specific areas of **matchIT**, please contact Support@helpIT.com.

The central feedback window will keep you informed of progress.

Viewing Matches

Once your file has been processed the *Matching Results* dialog is displayed, as shown below.



To look at the duplicates *matchIT* has found, choose *Verify Matches*. The first pair of duplicates is displayed, in the upper right corner: shown side by side:

First Record		Second Record	
UNIQUE_REF	0000000020	UNIQUE_REF	0000000109
ADDRESSEE	Mr W Summers	ADDRESSEE	Walter Somers
JOB_TITLE		JOB_TITLE	
COMPANY		COMPANY	
ADDRESS1	4 Ashburnham Close	ADDRESS1	4 Ashburnham Close
ADDRESS2	London	ADDRESS2	London
ADDRESS3		ADDRESS3	
ADDRESS4		ADDRESS4	
POSTCODE	N2 0NH	POSTCODE	N2 0NH
TELEPHONE	(081) 1231230	TELEPHONE	0181 1231230
PREMISE	4	PREMISE	4
CONTACT	Mr W Summers	CONTACT	Mr W Somers
SALUTATION	Dear Mr Summers	SALUTATION	Dear Mr Somers
MATCH_REF		MATCH_REF	

Matching Score: 80

This screen allows you to display each pair of potential duplicates in turn. The matching score is shown in the upper right corner, as well as to the left of each set in the list view. The matches are shown with the least likely dupes (lowest match scores) first, as you may only want to review the lower scoring pairs.

You can cut and paste between records, or simply verify whether the duplicates are candidates for flagging. Clearly, you don't have to wade through all the pairs of duplicates in this way, but *Verify Matches* is a good way of establishing the correct threshold above which it is safe to globally flag the duplicate records.

Some of the fields in the pairs of records are color-coded - this is to highlight where *matchIT* thinks the main differences lie between the two records. Those that are marked in red show fields which are clearly different in content, and those highlighted in yellow show where one field's contents are contained within the corresponding field in the other record.

The first pair displayed has a score of 80. This is because pairs of records that have scored less than 80 are not thought by *matchIT* to be duplicates. NB: The match score is not a percentage, but a grade to help separate out *true* matches from *false*. If there is a 'gray area' for the duplicates that have been found in a given file, with the default Matching Weights those duplicates will normally be in a score band of 80-85. Jack Whitson and Mr J Watson, both at the same address and postcode would be shown as a match scoring 80, when they could be different people.

However, this gives you the chance to go for marketing 'overkill' and flag one of these records, rather than risk sending someone two communications, one of which is wrongly addressed.

False Match

Flag Record

Find R

Match Set

Jump to Next Score

If any pair shown is not a true match, select the False Match button to unflag the match.

This button will flag the right hand record in the pairs view. The flagged record is then grayed out and the icon changes from Flag Record to Restore Record, which allows you to change your mind.

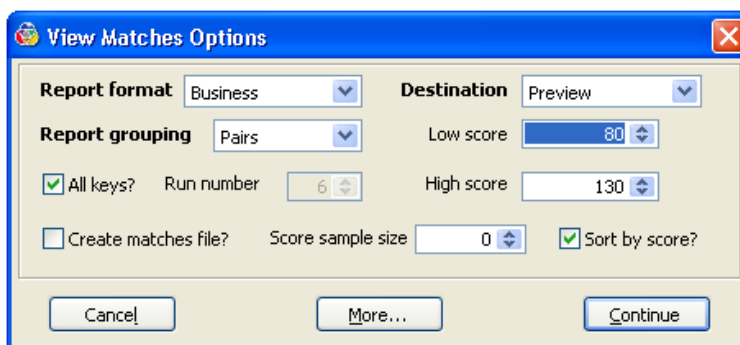
Use the Quick Find button to quickly find duplicate records that contain a certain piece of information.

Use the forward and backward arrow buttons to scroll through the pairs of duplicates.

There is also a Next Score button, to help you decide on the threshold score from which to flag duplicates globally.

If you right click on either record, you will see additional options for copying or combining data from the two records. Select *Done* when you have looked at the duplicates from each of the score bands.

Now select *View Matches* from the *Matching Results* dialog box. The following dialog is displayed:



This dialog is used to indicate which matches you want to print (or preview) and in what order. Change the Report grouping to *Sets* from the default *Pairs*, but leave the Report Format as *Business* with Destination set to *Preview*. Click on *Continue*.

A report is displayed listing all matching records, showing the unique references, name, address etc. You can click on any part of the report to zoom in or out, or use the scroll bars to see the different parts of the report.

On the print preview toolbar, click:



to print the report on the default printer and close it, or



to close page preview without printing.

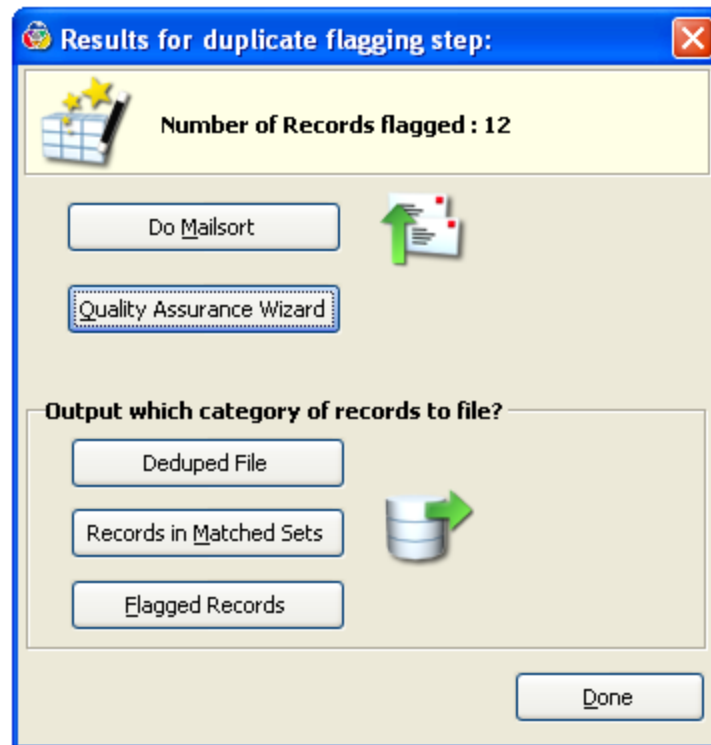
Note You can customise any of *matchIT*'s reports via the *Edit, Output Layouts* menu e.g. to show additional data items, or to change the formatting of the report. Please see *Going Further with matchIT* for further information.

Flagging Matches

Once back at the *Matching Results* dialog, select *Flag Matches* from the options at the bottom of the window. You are prompted to specify a score, equal to or above which it is safe to flag duplicates - so leaving the minimum score at 80 will flag all duplicates found. Once your flagging score is set, select the *Flag* button. A dialog will then be displayed showing the results of the flagging step.

The records flagged will be excluded from any deduped file output, but will be included when you want to export information about matching records. You can always undo the results of the Flagging Step by unflagging "flagged" records either individually via *Verify Matches* or globally via *Database Utilities*.

With a bit more experience using *matchIT* you will see how it is possible to flag all the closest duplicates automatically, and come back to deal with the less obvious ones interactively - perhaps after getting advice on how to decide which of a given pair to flag, from the client or department whose data it is you are processing. You can also use the *Intelligent Data Merge* option, which will ensure that any data present in the record being flagged, but missing from the record being kept, is copied over. This option is subject to rules that you can specify in the *Setup, Matching Setup* menu.

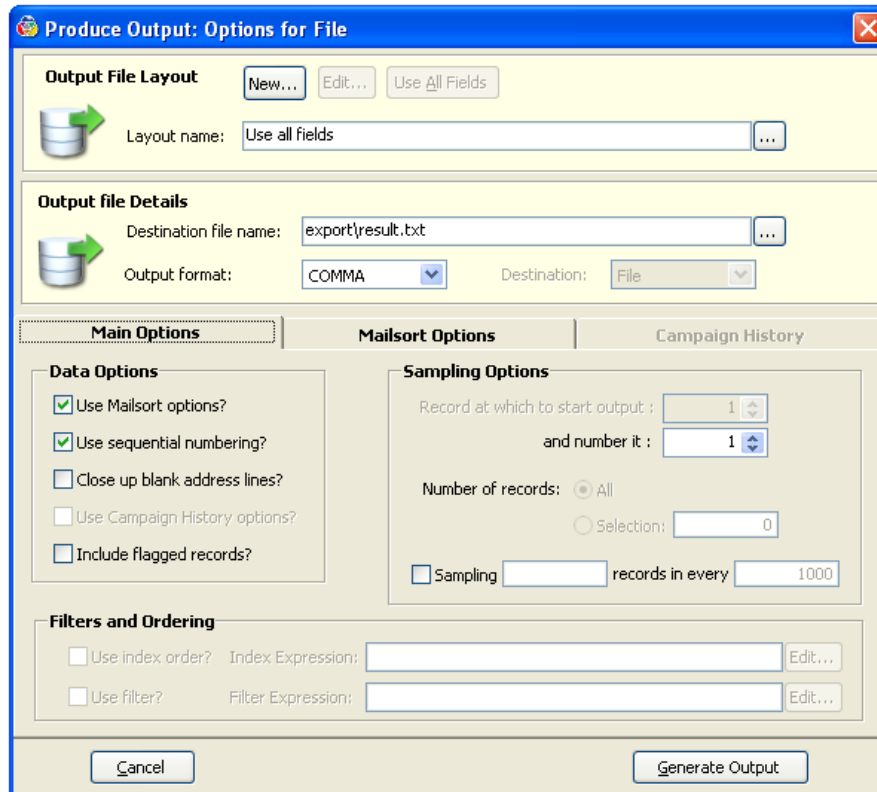


File Output

At this stage in the exercise, we have imported our data, found the duplicate entries and flagged them. Now we want to output a cleansed list. If this file is to be used for a "live" job, you should now select the Reports/QA Dashboard button from the screen above, to examine summary information about the file, "drill down" to view any suspect records, view data in different orders, output ranges or samples of the data.

If you choose *Records in Matched Sets*, *matchIT* will export unique reference numbers for matching pairs to a file. This will enable you to remove duplicate records from your source database and reassign "orphan" records, using a program external to *matchIT*. Choosing *Flagged Records* outputs only data for the records marked as "duplicates".

For this exercise, choose to output the *Deduped File*. This displays the *Produce Output* dialog which has 3 tabs, showing Main Options, Mailsort Options (only if you have the Mailsort module enabled) and Campaign History (only enabled in *matchIT* Campaign).



Produce Output: Options for File

Output File Layout New... Edit... Use All Fields

Layout name: Use all fields

Output file Details

Destination file name: export\result.txt

Output format: COMMA Destination: File

Main Options **Mailsort Options** **Campaign History**

Data Options

- ☒ Use Mailsort options?
- ☒ Use sequential numbering?
- ☐ Close up blank address lines?
- ☐ Use Campaign History options?
- ☐ Include flagged records?

Sampling Options

Record at which to start output : 1 and number it : 1

Number of records: ☒ All ☐ Selection: 0

☐ Sampling records in every 1000

Filters and Ordering

☐ Use index order? Index Expression: Edit...

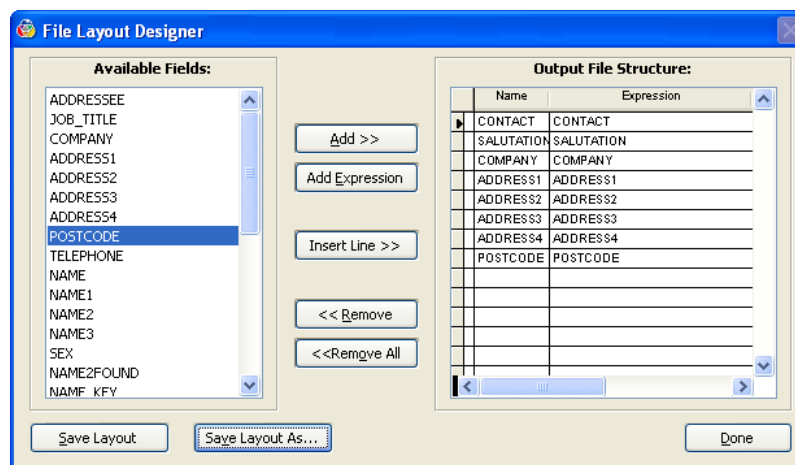
☐ Use filter? Filter Expression: Edit...

Cancel Generate Output

Output File Layout

matchIT allows you to save output layouts for future reuse. To create a new layout, select *New* in the *Output File Layout* section.

The resultant field picker (see below) shows you all fields that are available in the currently selected file (WORKFILE.DBF) in the left hand pane, and all the fields that have been selected for output in the right hand pane.



File Layout Designer

Available Fields:

- ADDRESSEE
- JOB_TITLE
- COMPANY
- ADDRESS1
- ADDRESS2
- ADDRESS3
- ADDRESS4
- POSTCODE
- TELEPHONE
- NAME
- NAME1
- NAME2
- NAME3
- SEX
- NAME2FOUND
- NAMF_KFY

Add >> Add Expression Insert Line >> << Remove << Remove All

Output File Structure:

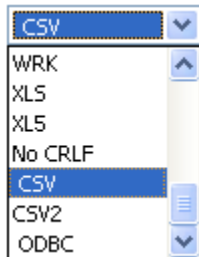
Name	Expression
CONTACT	CONTACT
SALUTATION	SALUTATION
COMPANY	COMPANY
ADDRESS1	ADDRESS1
ADDRESS2	ADDRESS2
ADDRESS3	ADDRESS3
ADDRESS4	ADDRESS4
POSTCODE	POSTCODE

Save Layout Save Layout As... Done

We could output the ADDRESSEE field from the original data, but when we asked *matchIT* to generate a SALUTATION in the Setup Wizard, we also generated a CONTACT field. Since this is a standardised name field that has been designed as the first line of the addressed item, this field should be the first that you select. It is one of

the last fields in the database, so scroll down the available fields list until you see it, and double click the field name to move it over to the right hand pane. Since this is a standardized name field that has been designed as the first line of the addressed item, this field should be the first that you select. It is one of the last fields in the database, so scroll down the available fields list until you see it, and double click the field name to move it over to the right hand pane. We will also select the SALUTATION field near the bottom of the list, because that is appropriate for the start of a personally addressed letter. Next select the COMPANY field, all the address lines and the POSTCODE field from the top of the available list.

Choose the *Save Layout* button and name it WORKFILE.OPL. Click *Done* to return to the *Main Options* dialog.



Although *matchIT* defaults to saving the file in its original format, we can change the output file format here. Drop down the Output Format list if you want to select a different format from COMMA - this is a comma delimited file without a header record, the same as the input file. If you want a header record to be inserted in the output file, to label each field, select CSV near the bottom of the drop down list. Other common options are Tab delimited, SDF (fixed width text), DBF and Excel (which is limited to 65,000 records). Microsoft Word can link to a comma delimited file as a data source for mail merge, so COMMA or CSV is a good choice for our exercise.

Next, click on the file selector button to the right of the *Destination File Name* to choose a different destination directory and name, if you wish and select *Generate Output*.

matchIT now generates the output file and (when finished) displays a message box that shows how many records have been output.

Conclusion

Now you have finished the whole process! Of course, this is a simple example, but *matchIT* is extremely flexible and sophisticated if you need it to be. As you become more experienced, you can fine tune *matchIT* to find all the duplicates in any data, no matter how badly structured or keyed. You can also find the common entries in multiple files (introduced in Exercise 2), automate frequent or complex jobs using the Job Script function of *matchIT* Pro and output address labels or mail merge pages direct to the printer.

Exercise 2

Multiple File Deduplication

What you have done so far is to dedupe one file. *matchIT* can also be used to find the common (overlapping) records across two databases, using the Merge/Purge functionality that is available in all versions except *matchIT Lite*.

The ability to find records common to two lists is a powerful function. It allows you to:

- *Purge* existing customers from bought-in mailing lists
- *Merge* databases from regional offices with a head office file, without creating duplication
- *Transfer Data* from matching records in one file to records in the other file
- *Write Overlapping Records* to output records that exist in both files to a third file.

All these options are available from the Merge/Purge menu. You can perform the Merge/Purge step on two databases of different structures e.g. one database may have the name all in one field, and the other split into title, first names and last name, as in this example.

If you have more than two files that you would like to *Merge/Purge* you can use the **Multiple File Wizard** (available in *matchIT Pro* and *matchIT Campaign* from the *matchIT Wizard* Pane). Alternatively, *matchIT* has the ability to work through files simultaneously in a process similar to that of Exercise 1. The Two File Wizard process is explained in the following sections.

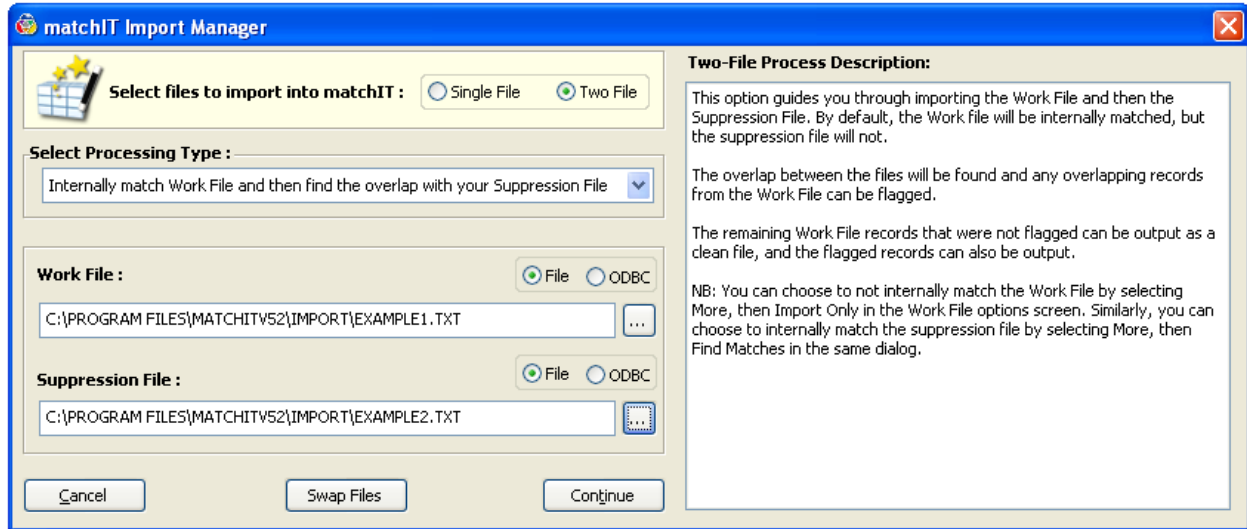
Starting the Two File Wizard



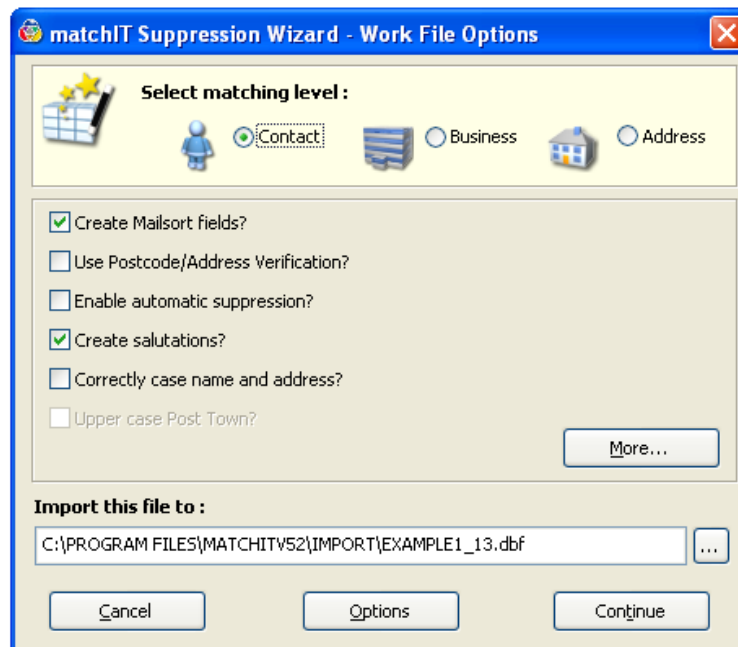
Select the Two File Wizard button on the left.

This will open a similar window to the single file process used in Exercise 1, but now you will select a second file and the process you would like to perform. Unlike EXAMPLE1.TXT, EXAMPLE2.TXT is in fixed width format, but the Setup Wizard will be able to recognise this.

For this exercise, we will *Find the overlap with your Suppression File*. Select that option from the Processing Type drop down. Next, open EXAMPLE1.TXT as your *Work File*. Then set your *Suppression File* to EXAMPLE2.TXT (EXAMPLE2 has no internal matches). Select Continue once you have chosen a process and the necessary files.



The next window displayed is the field-labelling window for EXAMPLE1.TXT, you may recognise this window as it is the same window used in Exercise 1. Make sure all fields are labeled correctly, and then select the *Continue* button.

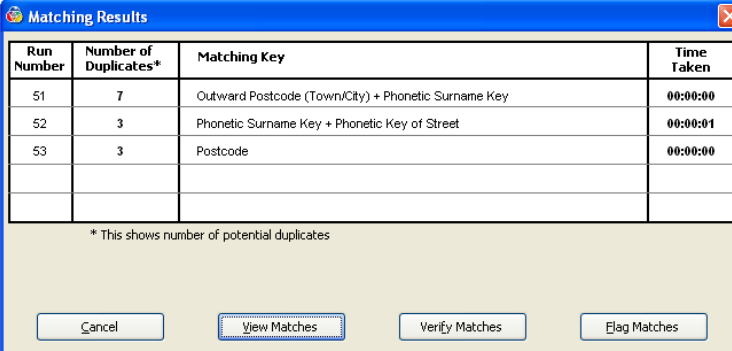


Depending on what you want to do with this file, you could select to generate salutations and/or case the data. You should select the matching level required i.e. whether you wish to dedupe to one record per Contact, Business or Address.

For this exercise, we are going to purge the records from EXAMPLE1.TXT that also exist in EXAMPLE2.TXT, so EXAMPLE2.TXT is a suppression or "stop" file. Tick the *Correctly case name & address* option, and set *matchIT* to a matching level of *Contact*. Click Generate Additional Fields, tick the generate salutation field option, and Save changes.

To make it clear which file is which, name this *matchIT* database WORKFILE.DBF by selecting the file selector button. Then select *Continue*.

Once matchIT has finished importing the data it will display the Matching Results window, see below. This is the same window that was displayed in Exercise 1. From here you can select to *View Matches*, *Verify Matches*, or *Flag Matches*. For this exercise, we are going to elect to *Flag Matches* for records scoring equal to or greater than 80.



Run Number	Number of Duplicates*	Matching Key	Time Taken
51	7	Outward Postcode (Town/City) + Phonetic Surname Key	00:00:00
52	3	Phonetic Surname Key + Phonetic Key of Street	00:00:01
53	3	Postcode	00:00:00

* This shows number of potential duplicates

Cancel View Matches Verify Matches Flag Matches

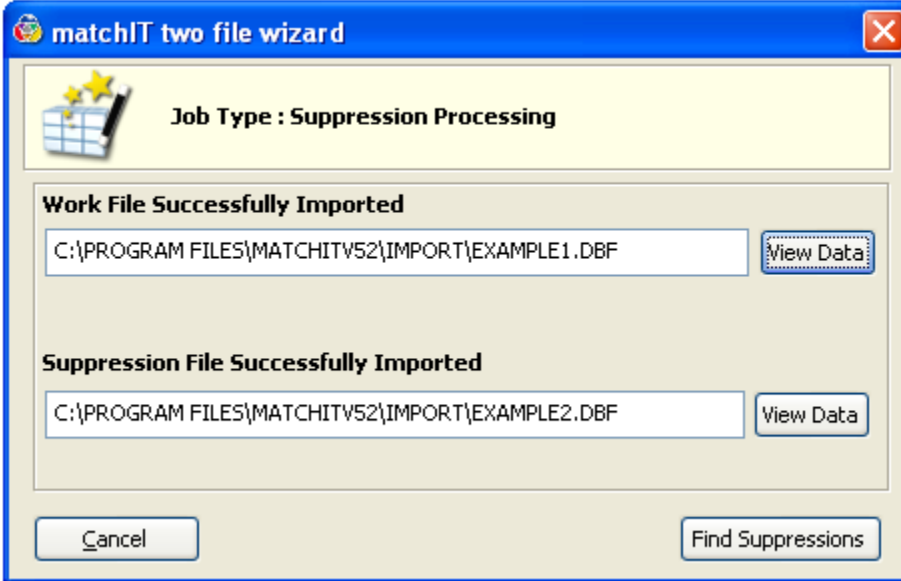
After EXAMPLE1.TXT has been imported and deduped, *matchIT* will begin to work with EXAMPLE2.TXT (the suppression file). *matchIT* will prompt the user to begin the import of EXAMPLE2.TXT, since this is a suppression file *matchIT* will not flag any duplicates within the file. Select *Continue* to begin importing EXAMPLE2.TXT.

EXAMPLE2.TXT will go through a similar process as EXAMPLE1.TXT, excluding internal deduplication. Save this imported file as STOPFILE.DBF and click *Continue*.

Now we are going to purge names from WORKFILE.DBF that already exist in STOPFILE.DBF. This is to ensure that we remove existing customers from the WORKFILE.DBF file that we just deduped.

Find Suppressions

To purge the STOPFILE records that also exist in the WORKFILE, select *Find Suppressions*.



matchIT two file wizard

Job Type : Suppression Processing

Work File Successfully Imported

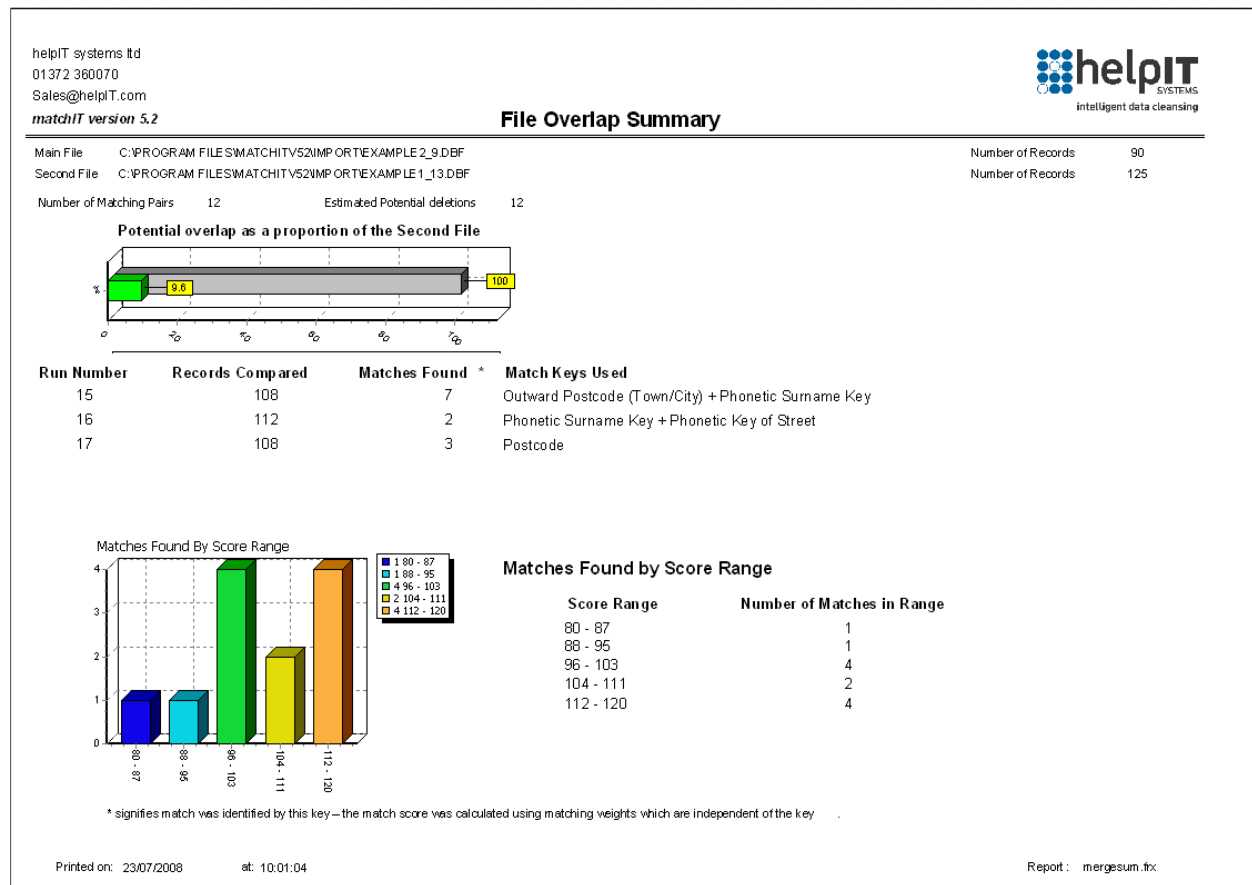
C:\PROGRAM FILES\MATCHITV52\IMPORT\EXAMPLE1.DBF View Data

Suppression File Successfully Imported

C:\PROGRAM FILES\MATCHITV52\IMPORT\EXAMPLE2.DBF View Data

Cancel Find Suppressions

The *File Overlap Summary* screen is then displayed:



This report shows the matching keys that found overlap, the score breakdown and the total number of overlapping records.

View/ Verify Overlap

After you close the *Overlap Summary*, the *Matching Results* window will be displayed. However, the *Matching Results* dialog now has different options from when you were deduping the single file:

Matching Results			
Run Number	Number of Duplicates*	Matching Key	Time Taken
15	7	Outward Postcode (Town/City) + Phonetic Surname Key	00:00:00
16	2	Phonetic Surname Key + Phonetic Key of Street	00:00:01
17	3	Postcode	00:00:00
* This shows number of potential duplicates			
<div> Cancel View Overlap Verify Overlap Remove Suppressions </div>			

View Overlap and *Verify Overlap* work in a similar way to *View Matches* and *Verify Matches*. Select *Verify Overlap* to see the matching pairs from the two databases:

Note how the databases' different structures are reflected in the layout of the screen. If there are more fields to see, you can use the central scroll bar to allow you to view the other fields. Choose *Done* to return to the *Matching Results* dialog when you have looked through the overlapping records.

Remove Suppressions

Having found the common entries between these two files, you can now select *Remove Suppressions*. You will then be prompted to *Flag* records that scored above a user-determined threshold, 80 is *matchIT's* default (40 for Address level matching). The *Results* window will open after the records have been flagged. From here you can utilize the *QA Dashboard*, output a *Cleaned File*, or output *Flagged Records*. Select *Cleaned File* to generate a clean file, one with no internal duplicates or overlapping records.

Note If at any stage you depart from the prompts for the usual options that the automatic dialogs display, you can select those options as required from the menus e.g. Output to File from the Output menu.

Generating a Clean List

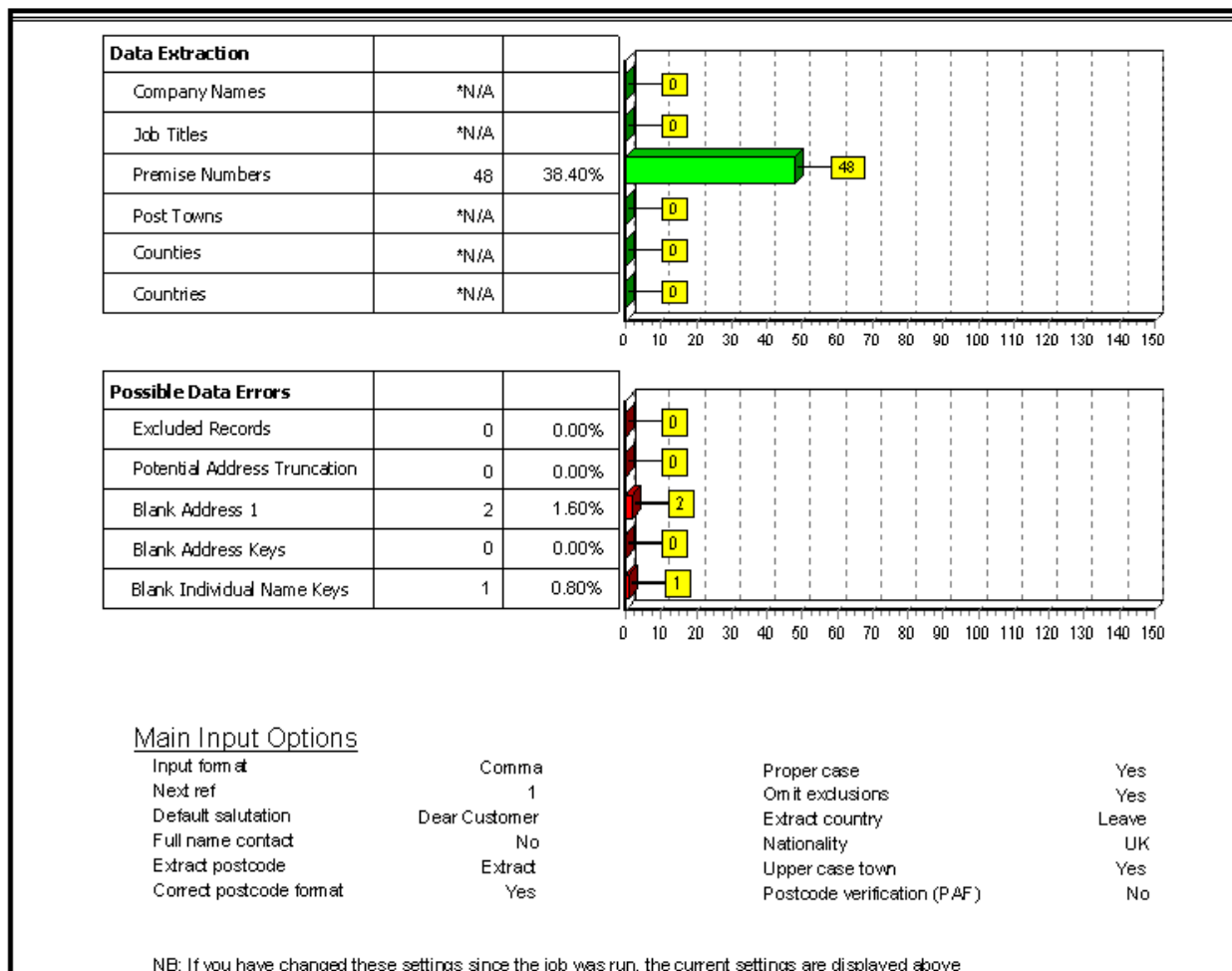
You can now generate a clean output file, one with no internal duplicates or overlapping records. After selecting the *Cleaned File* button, you will now see *matchIT's Produce Output* window. From this window you can determine the file layout, format and destination as you previously did in Exercise 1. After the file has been generated, you will see a window displaying the number of records output.

QA Dashboard

In the dialogs presented automatically after the Importation of Records, or the Deletion of Matches, you will see an option to utilize *matchIT's QA Dashboard*. The QA Dashboard is also available from the Output menu. Selecting this option results in the ability to display the Data Summary, View Records by Category and View Data within the DBF.

Select the Data Summary to view the same report as was generated in Exercise 1, but this time for the WORKFILE.DBF in Exercise 2. Scroll down or go to the next page to look at the information further into the report. This section of the report summarises information about:

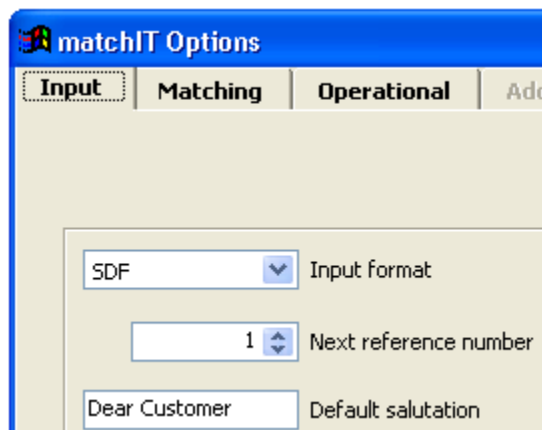
- Data Extraction
- Potential Data Errors, on records which should perhaps be excluded from the output
- Main Input Options



To "drill down", view various categories of records reported on the summary, close the report preview and select *View/Edit by Category* from the main QA Dashboard dialog.

The window displayed allows you to display all records meeting the selected criteria - most of them are categorised on the Data Summary.

You can select several categories to view at once. You can also choose to delete the selected records from this file before output, print them and/or write them out to a separate file. For this exercise, with **WORKFILE.DBF** open, tick *Default Salutation* and select *Browse on screen*. Select *Continue*, then *Continue* again on the next dialog to view any records for which *matchIT* could not derive a proper salutation e.g. records with no contact name, no prefixes with unisex first names and inconsistent first name and title e.g. Joanna Gifford Esq.



The *Default Salutation* shown is Dear Customer, but you can change this from the *Setup menu, Options, Input* tab, or from *More options* at the end of the Setup Wizard.

After you view the selected records, *matchIT* returns you to the View Records by Category screen to select more categories if you wish. To leave this screen, select Close.

Next, select View Data button (in the upper right) from the main QA Dashboard dialog. This displays the Browse Customer Database dialog. You can view *matchIT* databases in order of any field, click on *Order Records* then click on *Continue*. On the next dialog, scroll to the bottom of the field list and double click on *Salutation* as the field on which to order the new view.

Addresssee	Addresssee	Job_title	Company	Address1
THE MANAGING DIRECTOR	THE MANAGING DIRECTOR		SANFORD SERVICES	12 GREAT GEORGE
A WINTON	A WINTON	BUILDING MAINTENANCE	PROGRESSIVE OIL PLC	CASTLE IND PARK
M DEAN	M DEAN	PERSONNEL MANAGER	MANOR ENGINEERING	CASTLE INDUSTRIAL
J R MAGSON	J R MAGSON	PROD MGR	FRINTON WILKES LTD	EDMUND HOUSE
CHRIS PHILLIPS	CHRIS PHILLIPS		DECKCHAIR FABRICS	UNIT 23, PUDDLEWILL
JOANNA GIFFORD ESQ	JOANNA GIFFORD ESQ	OPERATIONS DIRECTOR	KEEN MAILING & MARKETING	MAILROOM
PAT O'DRISCOLL	PAT O'DRISCOLL		ROCKWADE PLASTICS LTD	P O BOX 263
PAT BROWN	PAT BROWN	MANAGING DIRECTOR	HANNAH DESIGNS	TRICENTRE ONE (3E
DR B H GEISLER	DR B H GEISLER	GENERAL MANAGER	ELK DESIGN (UK) LTD	PARADISE STREET
DR F MACPHERSON	DR F MACPHERSON	MARKETING MANAGER	M25 PUBLICATIONS	ALAMO HOUSE
DR S SAUNDERSON	DR S SAUNDERSON	NEW BUSINESS MANAGER	TRUTE COMMUNICATIONS	TRUTE HOUSE
DR J WATSON	DR J WATSON	MKTG DATABASE SERVS	BARTON MEDICAL GROUP	FOURTH AVENUE
FATHER ANDREW MACKAY	FATHER ANDREW MACKAY			ST FRANCIS CHURCH
MAJOR J L BELKSTEIN	MAJOR J L BELKSTEIN		DAVIS MANAGEMENT SERVICES LTD	33 LEYLAND AVENUE
MISS D BROAD	MISS D BROAD		PROGRESSIVE OIL PLC	CASTLE INDUSTRIAL
MISS F CALDERWELL	MISS F CALDERWELL	OFFICE MANAGER	BENEFITS AGENCY	AFFINITY HOUSE
MISS R K CUBITT	MISS R K CUBITT	MARKETING ASSISTANT	CAMBARD ENGINEERING	UNIT 2, ADAMS HOU
MISS L FOLKES	MISS L FOLKES	DATABASE SUPERVISOR	CUTTING EDGE DIRECT	PHOENIX MILL LOND
MISS A HEATH	MISS A HEATH	SALES MANAGER	GOVERNMENT LISTS UK	74 BERRY HILL
MISS STUART	MISS STUART		KMG MICROS	CASTLE INDUSTRIAL
MR & MRS DAVISON	MR & MRS DAVISON		AIMBISH LTD	REX HOUSE
MR J AIREY	MR J AIREY	MANAGING DIRECTOR	BAKERS DOZEN LTD	HOTSPUR HOUSE
RICHARD ANDERSON	RICHARD ANDERSON		KMG MICROS	CASTLE INDUSTRIAL
MR S BARR	MR S BARR	DEPUTY DIRECTOR	BUTTS & WORTH ENGINEERING	FAIRCHARM TRADING
MR RON BARTON LBIPP	MR RON BARTON LBIPP	MANAGING DIRECTOR	GIVENS ADVERTISING	227 QUEEN ALEXAND
MR T BENDON	MR T BENDON	PRINCIPAL	A.M.K. INC	ALLIANCE HOUSE
MR C G BICKNELL	MR C G BICKNELL		SCRIVENS MAIL GROUP	9-11 THE CENTRE
MR S BILTCLIFFE	MR S BILTCLIFFE	MARKETING MANAGER	FANDIGO WILLIAMS LTD	SHOWELL ROAD
MR F W BOWER	MR F W BOWER	MANAGING DIRECTOR	FRINTON WILKES GROUP	EDMUND HOUSE
TED BROWN	TED BROWN		MKL ELECTRONICS	ASH HOUSE
COLIN CAMPBELL	COLIN CAMPBELL		SILWOOD METALS	CASTLE INDUSTRIAL
MR G K CHILVERS	MR G K CHILVERS	LASER MANAGER	STANTON COPSHAW DIRECT	10 EAGLES WOOD BI
MR H COOPER	MR H COOPER	MANAGING DIRECTOR	AVERSONS MAILING	UNITS 9-10
MR A CUTHBERT	MR A CUTHBERT	STORES	PROGRESSIVE OIL PLC	CASTLE INDUSTRIAL
MR DAVID CUTTING-SMITH	MR DAVID CUTTING-SMITH	PARTNER	TRANSPORT DIGEST	FLOOR 3 NON-MARIN
MR CLIFTON DAVIES	MR CLIFTON DAVIES	OFFICE SERVICES CO-ORD	BRITISH PHILATELIC BUREAU	2ND FLOOR CITY HO
MR P DAWSON	MR P DAWSON	FINANCIAL CONTROLLER	BENTLEY JENNISON	55 DURAND ROAD
MARK DRIFFIELD	MARK DRIFFIELD	PROPRIETOR	CYCLES UNLIMITED	THE COACH HOUSE
MR R J EAST	MR R J EAST	MARKETING MANAGER	MOTOR TRADERS SUPPLIES	252 UPPER THIRD S
FELIX EASTWOOD	FELIX EASTWOOD	SALES MANAGER	UPSTAIRS MAILING	WOODROW DRIVE
MR G FILDES	MR G FILDES	INTERNATIONAL SVCS TEAM LEADER	DOONER INTERNATIONAL	UNIT 1, SADDLEHEAT
NIGEL FORSYTH	NIGEL FORSYTH		TAPS UK LTD	SHOBNALL COMPUT
MR B FRINTON	MR B FRINTON	FINANCIAL DIRECTOR	PROGRESSIVE OIL PLC	CASTLE INDUSTRIAL
MR LUKE GENTRY	MR LUKE GENTRY	SALES MANAGER	PLUMB BOOKING COMPANY LTD	14 PIERPOINT STRE
MR GREVILLE GIVENS	MR GREVILLE GIVENS		GIVENS ADVERTISING	227 QUEEN ALEXAND

In the data window shown, scroll towards the right. *matchIT* has added several fields after the last field from the original data - these are the phonetic and other key fields used by *matchIT* to search for matches. At the end, you can see the generated *Contact* and *Salutation* fields. *matchIT* has derived the *Contact* field from the supplied name, and worked out the correct *Salutation*. By viewing the data in order of a field, you are more likely to see unusual and perhaps suspect values of that field near the top of the data, which is a very useful *Quality Assurance* technique.

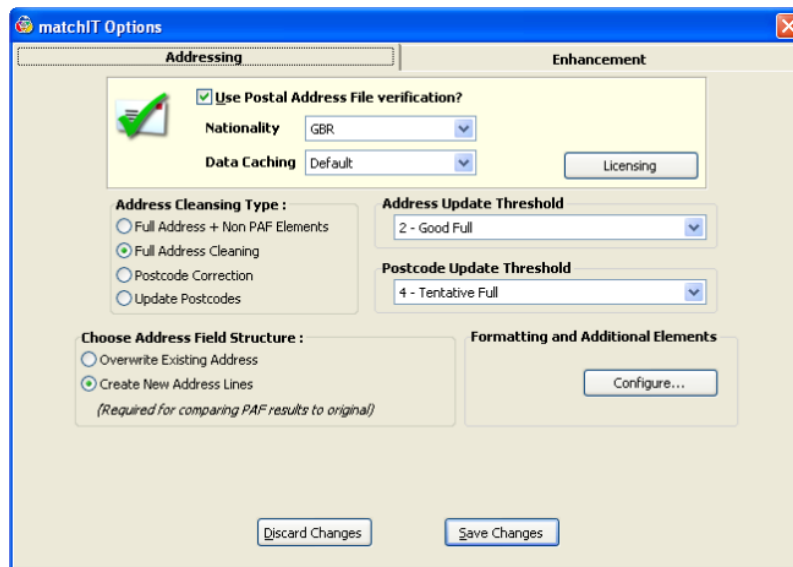
If you scroll down the list, and compare what's in these fields against the input name fields, you will see how *matchIT* deals with complex and uncommon name structures, even when an input file has the contact name within one field.

Press Escape to close the view, and then choose the Output tab from the main QA Dashboard dialog. If you chose Mailsort previously, you will see a Mailsort tab, otherwise the Output section will allow you to output 1 in N samples and selections of records. Select *Close*

Postcoding and Address Validation

Any version of *matchIT* can have the UK Postcoding and Address validation module, *addressIT*, integrated with it, using the addressing engine supplied by GB. The standard evaluation configuration does not include *addressIT*, but you can evaluate it if you request it - we will then send you a CD with the Royal Mail Postal Address File (PAF).

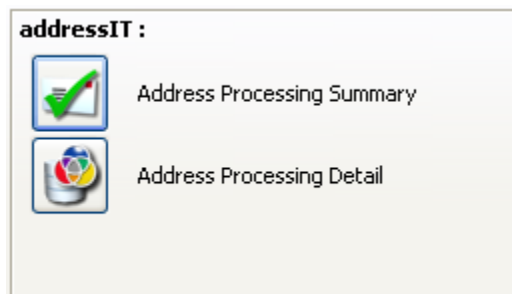
As with all of *matchIT*'s operations, you have to first import your data into a DBF file. You can use the Single or Two File Wizards on the left pane to do this. Once your data file has been imported via the wizard you need to tick the *Use Postcode/Address Verification* option in the Processing Options window. You should use the same rules for naming fields as are described in Single File Wizard Operation on page 7. You will then be able to select the *Addressing Options* screen. The key options here are described below.



- *Address Cleansing Type* - controls how the address and postcode are updated:
 - *Full Address + Non PAF Elements* - retains input address elements that are not on PAF e.g. customer preferred elements.
 - *Full Address Cleaning* - doesn't retain input address elements that are not on PAF.
 - *Postcode Correction* - verifies and corrects postcodes, but does not update addresses to those held on PAF.
 - *Address Update Threshold* - This is used to select the success threshold level that a record must reach before any changes are made to its address. As well as being named, the different levels have been numbered from 1 to 5. 1 ('Verified Correct') is the most successful level that a record can achieve when matching against a PAF record; 5 ('Unmatched') is the least successful.
 - *Postcode Update Threshold* - This option works in the same way as the Address Update Threshold, however it relates only to changes made to postcodes. This can be (and is by default) set to a slightly lower level than the Address Update Threshold. The reason for this being there is generally less of the address that needs to match a PAF file record in order to determine that addresses correct postcode.
 - *Choose Address Field Structure* - allows you to overwrite the incoming address with the address from the PAF or to create new address fields to store the PAF address.

- *Formatting and Additional Elements - Configure* - allows you to add additional address elements such as County and DPS and choose address formatting options.

After processing, you can review the postcoding results from the QA Dashboard (when prompted or selected from the Output menu).



Address Processing Summary provides a summary report of the postcoding success rate.

Address Processing Detail "drills down" to the records that have been changed or could not be postcoded. From this latter screen, you can select which categories of matches to review and either view multiple records at once. If you created new address fields to store the PAF address, you can also choose the Compare button to see the full details of the new address and postcode together with the old address and postcode in a side by side view. You can then choose to revert to the old address and postcode for any record.

Suppression/ Enhancement Service

This service is accessible from the Jobs menu using the Suppression/Enhancement option, and also from the very end of the processing wizard. Through this option even if you don't have suppressIT and addressIT, you may now:



- identify suppressions (eg deceased records and people that have moved house - known as goneaways)
- Append additional data via consumer names lookups
- Append additional business data (eg SIC Codes, phone numbers, company registration numbers etc)

If your mainfile contains consumer records (ie it does not contain a company field), then you will see the following screen when selecting the service.

From this screen you may select any suppressions you would like to run using the suppression options (choose advanced options if you want to specify the suppression file provider (eg Equifax and the ReAD Group etc. You can also enter your predicted production and mailing costs to get a ballpark cost saving that you might see by using the service (and hence removing unwanted mailing records from your database).

Additionally, you can click on the enhancement tab to choose from various different data enhancements that you might want to make to your data. If your mainfile contains business data, then only a business enhancements screen will be available.

When you have selected the options you want, click continue and you'll see a message indicating a receipt file has been written to your computer. You will then be taken to a user details screen where you must enter your contact details. From there click continue to have a copy of your data file encrypted and uploaded to helpIT.

When you receive your processed data back from helpIT, you will need to integrate these results into your original mainfile. This is done by copying the return file into the same folder that your mainfile resides in, and then opening your mainfile and choosing 'Remove Bureau Suppression/Update Enhanced Data' from the jobs menu. This will flag any records that should be suppressed any add any additional data that you've requested to the mainfile.

Further Information

On-Line help is available from the F1 key. The Getting Started Guide, Frequently Asked Questions and full User Manual are available from the Help menu in *matchIT*. If you have any problems or questions, or want to know if the results that you are receiving can be improved, you can refer to the Frequently Asked Questions section of our web site. Alternatively, please don't hesitate to call us. With our experience, we should be able to help you get the best out of *matchIT* very quickly and we'll be only too pleased to help.

If you would like information about the core standardization and matching components of *matchIT*, for plugging into your own systems, please contact us for information about the *matchIT API*.

For help and more information, please email or visit:

Sales@helpIT.com

www.helpIT.com

Other contact details are as follows:

UK Office

helpIT systems ltd
Swan House
24 Bridge Street
Leatherhead
Surrey KT22 8BX

US Office

helpIT systems inc.
51 Bedford Rd
Ste 9
Katonah
NY 10536

Tel:

+44 1372 360070

(866) 332-7132

Fax:

+44 1372 360081

914 232 1429

Support email:

Support@helpIT.com

Support.US@helpIT.com