

DbCapture user and administrator guide

Document Management & Imaging



This document describes the DbCapture software. Part of the Version One document management suite – award-winning paperless office solutions.

Version One Ltd

Pentland House, Village Way
Wilmslow, Cheshire SK9 2GH
United Kingdom

t: +44 (0)1625 856500

e: info@versionone.co.uk

www.versionone.co.uk

Last edit: 20 September 2011

Copyright and ownership

© 2011, Version One Limited. All rights reserved.

Version One software is supplied under licence and may be used or copied only in accordance with the licence terms. The presence of information in this document does not entitle you to access or use the software described.

Accuracy and updates

We tried to ensure that the information in this document is correct, but cannot accept liability for any errors and omissions. Version One software products are being developed continuously, so the published information may not reflect your own system. In addition, your own software may not contain all the features described, depending on the packages installed and their configuration. This document was written alongside v2.1 of the document management suite.

No warranty, either expressly or implied, is made by anything in this document, including but not limited to any relating to merchantability or fitness for any particular purpose.

Document feedback

We welcome your feedback on this document, please email any comments or queries to: documents@versionone.co.uk

Confidentiality

This document is confidential. No part of this document may be reproduced or transmitted in any way without prior written permission of Version One.

Ownership

Version One retains ownership of this document. Approved third parties may have reused and reworked the content, but original ownership lies with Version One and must be credited in any reworking. Sharing this document with approved re-sellers does not imply any sharing of the software ownership, source code, look and feel, etc, and does not override any license agreement.

Trademarks

All trademarks are the property of their respective owners and are used with their thanks.

In this document...

Introduction	1
Version One document management suite	2
Document creation and distribution	2
Document scanning and archiving	3
Document data capture	4
Document authorisation	4
The document management suite applications and client tools	5
See also:	6
Using the document management suite	7
Before you log in	7
The login window	7
Programs menu option	7
Recently used menu option	8
Desktop shortcut option	8
Customising windows	8
Toolbars	8
Using the windows	9
Sizing columns	9
Sorting columns	9
Moving columns	9
Filtering columns with search	9
Clearing your search criteria	10
Date picker	10
Complex searches	10
Clearing multiple search criteria	10
Choosing fields	10
Using help	12
Links and how they work	12
Screen shots	12
The icons	12
Finding a help topic	13
Glossary	14
Accessing the applications – logging in	15
DbLogin admin console	16
Securing your data	17
Server	17

Scan station	17
About us	18
Getting in touch	18
Problems and queries	18
Copyright and small print	18
Help about	19
Standard toolbar options	19
DbCapture	21
About DbCapture	23
Purchasing cycle	23
Sales cycle	23
Capture queue	23
Quality of originals	23
About DbCapture export	24
About DbCapture admin client	26
Filtering the display	26
Processing documents	26
Toolbar	28
Sub-queues	29
DbCapture admin context menu	30
About DbCapture interactive client	31
Capture problems.....	32
Validating data.....	35
Capture templates	36
Installing templates.....	36
Editing templates	36
Understanding document identifiers	37
What is a document identifier?.....	37
Using document identifiers	37
Creating templates	37
Special cases	38
Summary	39
Creating and using anchors	40

What is an anchor?	40
What is a floating anchor?	40
What makes a good floating anchor?	40
Creating floating anchors	40
Line level capture	44
Using line level capture	45
Troubleshooting	47
Capture properties	52
General tab	52
Line level tab	53
PDF import	54
PDF import module	57
Technical outline	57
Using the PDF module	58
How it works	58
Installing and configuring	59
Installation	59
Configuration	59
DbCapturePDFFileWatcher	59
DbCapUtil command line utility	59
Document background processing	60
OCRFromMaster (.dbcs [main] section)	60
UnpackedFormat (.dbcs [main] section)	60
Passing document type to DbCapture scripts	61
PDF importing FAQs	62
Glossary	63
Index	66

Introduction

Version One document management suite

Overview: the document management suite provides scanning and document management facilities for electronic handling of paper documents. It can be used to automate and simplify your:

- sales document cycle (order to cash)
- purchasing cycle (purchase to pay)
- other archiving requirements (HR documents, technical documents, etc).

These solutions save significant amounts of time, money and storage space. Typical documents:

- sales order invoices
- purchase orders
- statements.



Software installation is normally handled by Version One, or another specialist, and is not covered in this help.

The Version One document management suite is a comprehensive collection of software tools to automate document processes within their core application environments. Using the software you can replace paper-based processes with electronic procedures, eliminating printing, posting and manual filing.

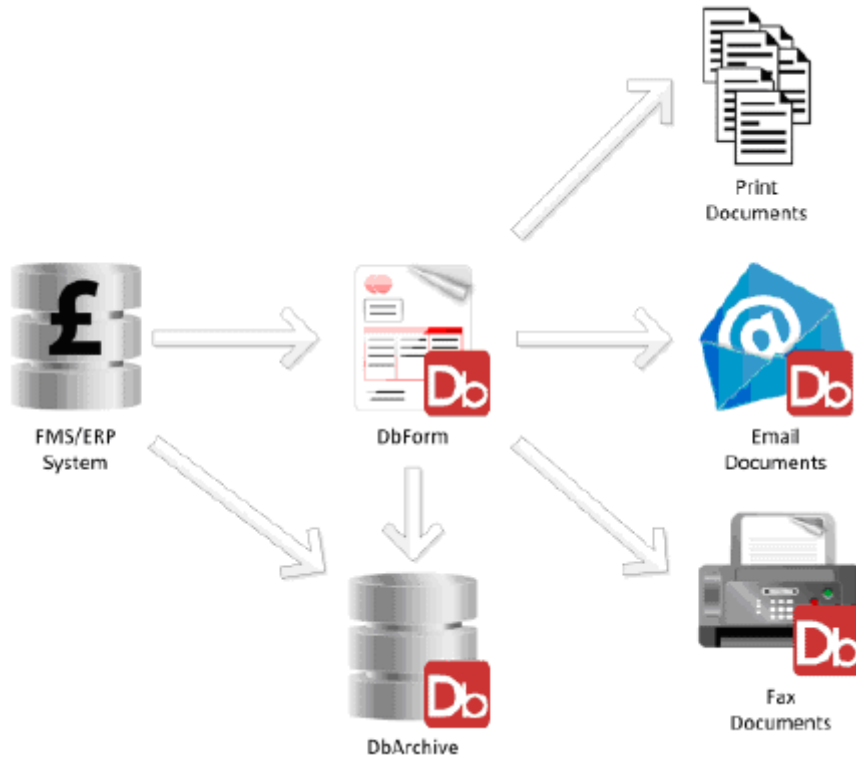


There are many different ways of configuring and using the software, using different combinations of the core product set. You may not have the complete applications suite installed.

The document management suite is made up of a number of integrated application modules that work together to give the desired functionality. These applications are broken down into the following areas:

Document creation and distribution

Documents based on data from core systems may be formatted as professional output documents which can then be printed, faxed, emailed and archived based on defined business rules.



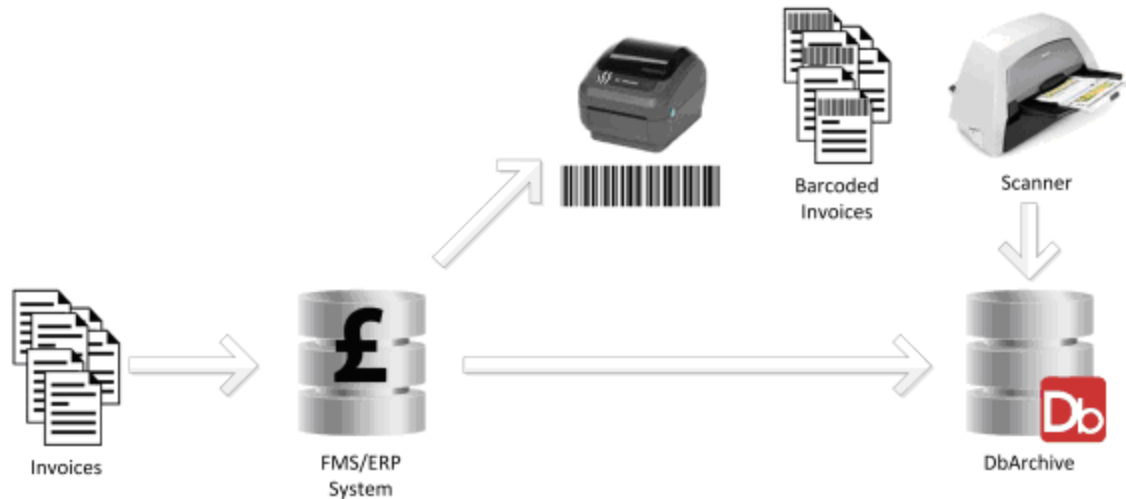
Several applications support these processes:

Application Description

DbForm	Document formatting and business rules management.
DbFax	Sending documents by fax.
DbMail	Sending documents by email.
DbArchive	Archiving outbound documents.

Document scanning and archiving

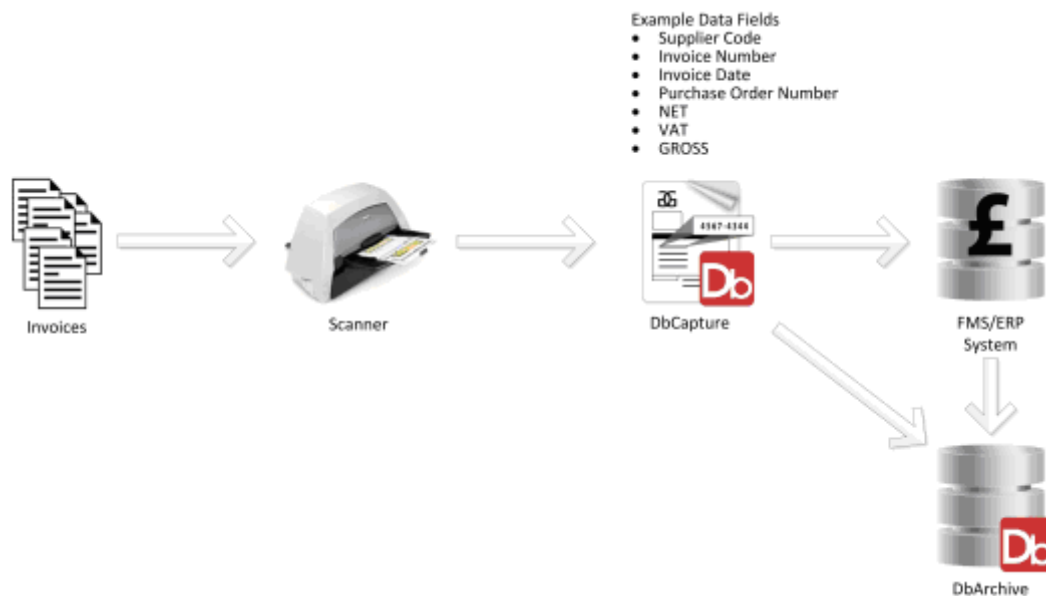
Incoming paper documents such as invoices and orders may be scanned and stored for easy access and for statutory purposes (eg to comply with UK HMRC legislation). The basic process makes use of barcoding techniques for document separation and recognition.



The application supporting document scanning and retrieval is DbArchive.

Document data capture

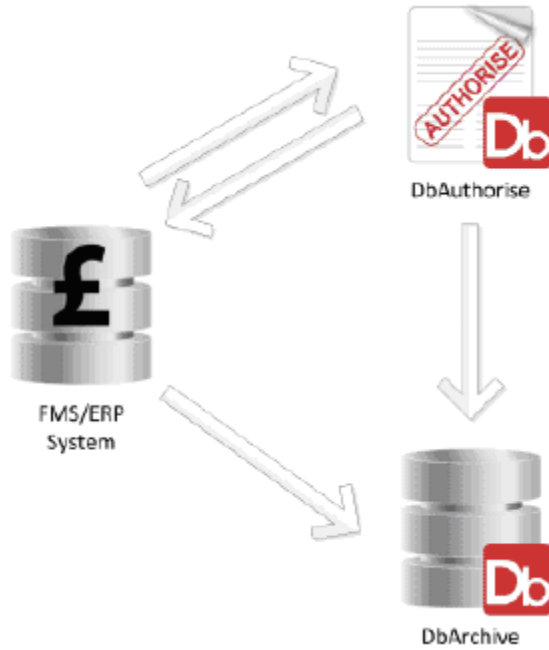
This process typically involves scanned documents such as purchase invoices. The document data can be intelligently captured using OCR technology and validated, so transactions can be automatically posted to the core finance system.



The application supporting intelligent data capture and transaction posting is [DbCapture](#).

Document authorisation

This process typically involves documents such as purchase invoices captured using DbCapture and now requiring coding and approval for payment.



The application supporting document routing and authorisation is DbAuthorise.

The document management suite applications and client tools

The following applications and associated client tools are available as part of the Version One Document Management suite:

Application	Client tool	Description
DbLogin	DbLogin admin console	The master utility for managing the Version One Document Management suite and who has access to it.
	Green meter console	Shows the environmental benefits from using the suite.
DbForm	DbForm GUI designer	Design the format of the outgoing documents.
DbFax	DbFax client	Manage the DbFax system. Create and send faxes from a client PC.
DbMail	DbMail admin client	Manage the DbMail system.
DbArchive	DbConsole	Configure the Tables and Access Levels available within the DbArchive system. Manage the DbArchive system.
	BLOB type manager	Configure new document formats on DbArchive systems using the Native Documents Option.
	Failed deposits manager	Deal with documents that have failed to deposit correctly into DbArchive.
	DbQuery	Document search and viewing facilities based on the user's designated access rights.

	DbWebQuery	The web based alternative to DbQuery for document search and retrieval.
	DbWebRetrieve	Additional functionality to DbWebQuery, providing better searching and enhanced usability.
	Index manager	Amend the search data values for stored documents and copy or move documents between DbArchive tables.
	DbDeposit	Store ad-hoc documents, which exist already directly into the archive.
	DbScanner	Scan documents for storage in DbArchive and optional intelligent data capture within DbCapture.
DbArchive export	Export documents from DbScanner into DbArchive	
DbCapture	DbCapture export	Export documents from DbScanner into DbCapture.
	DbCapture admin	Manage the queue of documents being processed within DbCapture.
	DbCapture interactive client	Process individual documents within DbCapture and to create templates for new document formats.
DbAuthorise	DbAuthorise client	Coding and approval of documents within DbAuthorise. Manage the Routing Rules and other management tasks.



Customisation: Some applications may not be installed on your system, depending on your license, any packages you are interfacing with (eg Accounts) and choices made during installation. You may have modules embedded in a parent application, so you would not necessarily see them branded as Version One.

See also:

[Using the Version One document management suite](#) more on using the applications
[Using help](#) more about using this help
[Using the windows](#) general conventions for working with data in the software.

Using the document management suite

Overview: the Version One document management suite has a number of common features. This topic outlines some of the common functions used across the product family.

Before you log in

To log in to the Version One document management suite you need the following:

- the software must be installed on your PC or laptop
- user name — your system administrator should have created this for you. This is your identity on the Version One document management suite
- password — your system administrator should have created this. If you are not using Windows authentication you can change this once you log in to the Version One document management suite.



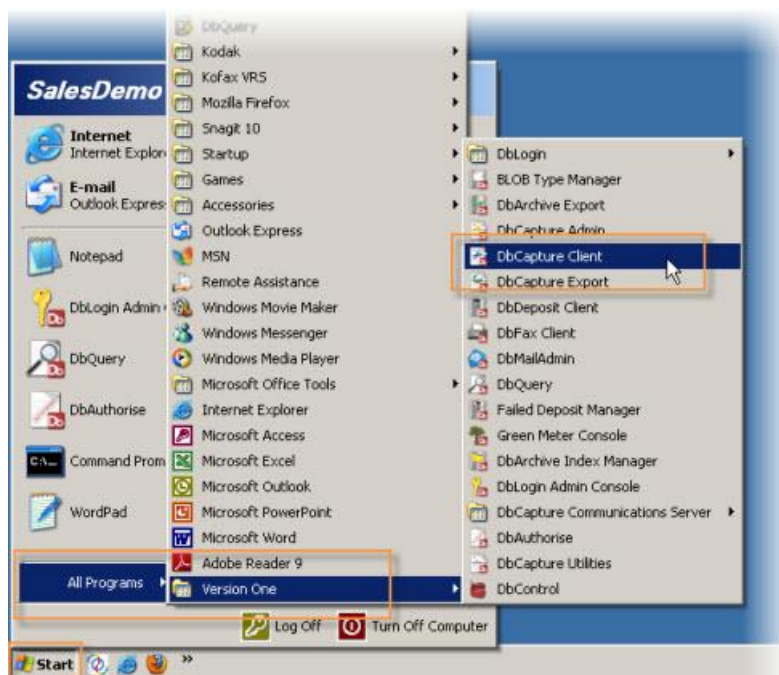
Customisation: your system may be using Windows authentication to log in to the document management suite automatically. If so, the [DbLogin window](#) is not displayed. Using Windows authentication allows you to use your regular PC logon to access the software without an additional login.

The login window

Once the Version One document management suite has been installed, there are several ways for you to display the [DbLogin window](#).

Programs menu option

Version One is displayed in the list of programs on your computer. Click **Start > All Programs > Version One > [Program Name]**. For example:



Recently used menu option

Once you have used the Version One document management suite it may be displayed in your Recently Used menu above the Start button. You can also use this to start the program.



Desktop shortcut option

You can drag and drop the link to your Version One document management suite program from one of the menu options onto your desktop. Click the mouse, hold the button down, drag to the desktop and release the mouse button.



You can then click this desktop shortcut to start the program.

Customising windows

Many of the windows in the document management suite can be configured to suit your preferences, eg:

- Dragging and dropping columns to change the order they are displayed
- Hiding some columns
- Sorting columns by various filters.



Customisation: once a screen has been changed, the system uses the new layout until you reset it to the original layout. Example screenshots shown in this help therefore might not exactly match what you see on your system.

Toolbars

There are a number of common icons used throughout the system. As you become familiar with the toolbar icons and their use in different contexts, their use should become intuitive. The common Help menu functions like change user, environment, language and password are described in [Help About](#).



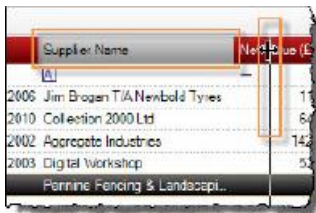
The options available depend on your particular document management suite configuration.

Using the windows

In many places within the software you can customise the display to change the way document results are shown. For example sizing, sorting and moving columns. You can also filter columns with search, run complex searches across columns, use the date picker and choose the fields you want to display.

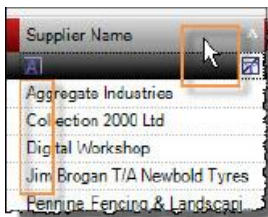
Sizing columns

You can change the size of columns by clicking and dragging. Here we are making Supplier Name wider to show the full name.



Sorting columns

You can sort the data by clicking on a column heading. Here clicking the Supplier Name column shows documents sorted alphabetically by supplier.



Click the column header again to sort the column in reverse order.

Moving columns

You can customise the column layout by clicking and dragging the column header to a new location. Here we are dragging Supplier Name to the left.



Filtering columns with search

Many columns in the software have a search filter used to narrow the results shown. Here we are filtering Supplier Name to show suppliers beginning with 'agg'.



Clearing your search criteria

Clear the search criteria by clicking the small cross alongside the search term:



Date picker

Use the date picker to search for documents for a particular date:



Complex searches

More complex search queries are available for some fields. Click the drop-down arrow to choose a search type and add your criteria:



Clearing multiple search criteria


Some windows allow you to clear all your search criteria at once by clicking a button:



Choosing fields

Some windows allow you to choose which fields to display. Click the icon to show the available fields and mark the ones you want with a cross:




 Try out the layout options to get the display how you want it. You are only changing the way information is shown, not changing the data. Some windows can't be customised.

Using help

When you first open help the introduction page is shown, along with navigation tabs (Contents, Index and Search) to find more topics.

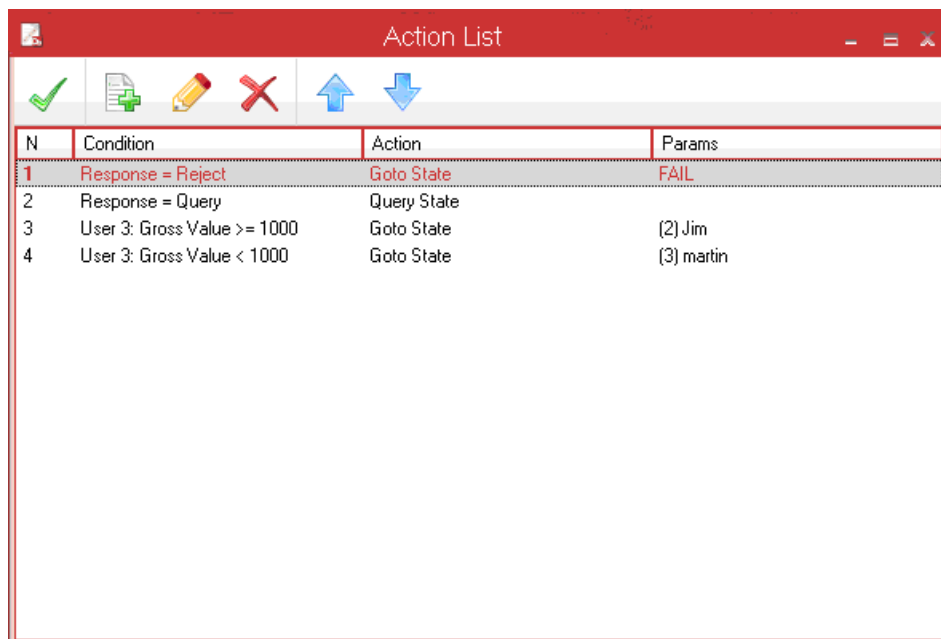
Links and how they work

Throughout help, you will see links such as [Links and how they work](#). Holding the mouse over a link changes the cursor to a pointer . Click for one of the following:


- more text
- new window opens (click anywhere to close)
- another help topic (use the Back button to return to the original topic).

Screen shots

Sample screen shots are shown where they help clarify the topic. To save space and keep the help briefer, they are often shown in drop-down links like this: [Show/hide screen](#). Click the link to **show the graphic**. Click again to **hide the graphic**.



N	Condition	Action	Params
1	Response = Reject	Goto State	FAIL
2	Response = Query	Query State	
3	User 3: Gross Value >= 1000	Goto State	(2) Jim
4	User 3: Gross Value < 1000	Goto State	(3) martin

 Some screenshots are mock-ups of real data, made anonymous by blurring key information. Any resemblance to real companies does not imply an endorsement of the software or an active trading relationship between companies.

The icons

These icons help you use the help more effectively.

Icon	Description
------	-------------



Hide

Hide the navigation and glossary panel.



Show the navigation and glossary panel.



Go back through the topics you have looked at.



Go forward through the topics you have looked at.



Print the current topic.



Finding a help topic

You can find the help you need several ways:

- contents
- index
- searching the text.

Search the contents

Searches the **Contents** and **Topics** (like chapters and sections in a book):

1. Click Contents.
2. Click one of the Books 
3. Click a Topic  to show the content on the right.

Search the index

Searches the predefined entries in the **Index** (like the index of a book):

1. Click **Index**.
2. Type your search term.
3. Click an index word to show the topic.



Help does a best match against your term. For example: if the index contains both 'Hidden' and 'Hiding' and you search for 'Hid', 'Hidden' is highlighted as the first word containing that string. If you type 'Hidi', 'Hiding' is highlighted ('Hidden' no longer matches the string). Try synonyms for what you are looking for, eg 'house', 'dwelling', 'residence'.

Search the text

This method searches all the text in the help (like searching the text of a book):

1. Click **Search**.

2. Type the word you are searching for.
3. Click **Go**.
4. All topics containing the search word are listed.
5. Click the topic you want.

Glossary

Some of the terms used in the help are defined in brief glossary entries. These are self-contained descriptions and don't link anywhere else.

Accessing the applications — logging in

Overview: a login window is displayed when you click a Document Management suite icon. The same style of login window is used for all applications. This is where you enter your login username and password.



Field	Description								
Username	Enter your user name. This may be your Version One user name or your Windows user name, depending on the option you select in the Log on using field.								
Password	Enter the Version One or Windows password for the user name you entered.								
Log on using	Select how you want to log on to the Version One Document Management suite:								
	<table border="1"> <thead> <tr> <th>Option</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Windows</td> <td>Windows Authentication attempts an Active Directory (AD) User Password login — specify your Windows password to validate against Active Directory.</td> </tr> <tr> <td>Internal DbLogin</td> <td>Version One document management suite Login ID and password — attempts an internal login for the user name and password entered.</td> </tr> <tr> <td>Current Windows user</td> <td>Uses the Windows User currently logged in to this computer. Attempts an AD Login, replicating the default behaviour on loading</td> </tr> </tbody> </table>	Option	Description	Windows	Windows Authentication attempts an Active Directory (AD) User Password login — specify your Windows password to validate against Active Directory.	Internal DbLogin	Version One document management suite Login ID and password — attempts an internal login for the user name and password entered.	Current Windows user	Uses the Windows User currently logged in to this computer. Attempts an AD Login, replicating the default behaviour on loading
Option	Description								
Windows	Windows Authentication attempts an Active Directory (AD) User Password login — specify your Windows password to validate against Active Directory.								
Internal DbLogin	Version One document management suite Login ID and password — attempts an internal login for the user name and password entered.								
Current Windows user	Uses the Windows User currently logged in to this computer. Attempts an AD Login, replicating the default behaviour on loading								

a client application. Doesn't require a password.

Validation of Windows login with Active Directory (AD) uses a secure socket (ssl) connection from the DbLogin server to Active Directory. If this fails, Windows login is not possible (other than the basic AD login as the current Windows user). AD user verification from the server is only supported from the Windows DbLogin server, not the Unix versions.



The application version number is shown for information — you'll need this information if you contact your support team with a query.

Button Description

OK Log in using the details entered.

Cancel Exit without logging in.

After validating the login any DbLogin password rules are applied, eg prompting you to update your password. After a successful login, DbLogin applies the default environment (normally 'Live', unless configured otherwise) and checks:

- the user has access to their default environment
- the application being loaded is present in the default environment
- valid CAL access can be granted.



Any failure prevents access to the client application.

DbLogin admin console

Login administration is covered in a separate topic for **system administrators only**.

Securing your data

Overview: you need to make sure that your data is securely backed up in case of problems. A secure backup policy is crucial to any IT system. The details of your backup policy is beyond the scope of this help, and not covered by the document management suite or Version One project management. These notes are for general guidance only.



This topic is for **system administrators only**.

The document management suite is a client-server application with components running on both a central server and client PCs.

Server

Back up the entire folder:

```
d:\versionone\
```

Scan station

If you have the ability to clone a machine, make an image of the scan station. Otherwise you should backup:

```
c:\versionone\
```

```
c:\Program Files\Version One
```



These folder names assume your software has been installed in the default locations. Backup up your own installation folders if different from those above.

About us

Getting in touch

This software is from Version One Ltd. If you have any comments on this help or suggestions for improvements, please get in touch. You can email info@versionone.co.uk. We're happy to hear from you.

Problems and queries

If you have a query on using the software you should get in touch with your support contact. In many cases your software comes from a third-party, who is normally your first point of contact for queries.



Pentland House, Village Way, Wilmslow
Cheshire SK9 2GH UK

Tel: +44 (0)1625 856500
Fax: +44 (0)1625 856501

www.versionone.co.uk

Copyright and small print

© 2011, Version One Ltd. All rights reserved.

All trademarks are the property of their respective owners and are used with their thanks. All rights reserved. This help was written alongside v2.1.0 of the document management suite. Last revised 19 August 2011.

Help about

Overview: each of the applications has a Help About window displaying information about the version of the application you are using. Click Help > About [application name], eg Help > About DbQuery to display the About DbQuery window.



You will need the version number if you contact your support team with a query.

Click OK to close the window.

Standard toolbar options

Many of the applications have standard toolbar buttons with frequently-used functions. The same functions are also on the Help menu:

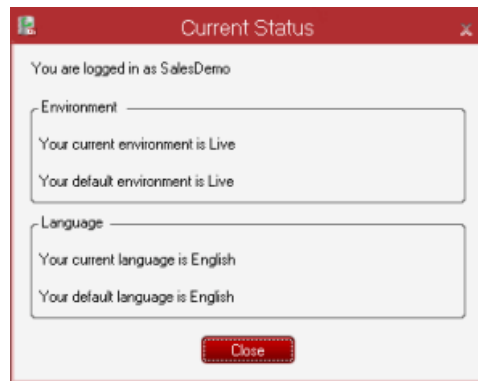


Shows the Help About window.



Use the drop-down to change your system configuration.

Option	Notes
Current status	Use this to check the current status of the environment and language settings.



Change environment

You can have multiple environments, eg Live and Test. Choose the environment from the drop-down box. You can also choose to make this environment the default using the checkbox.



Change user

Change the current user to a different login. Add the username and password for the new user in the Change User window.



Change password

Change the login password for a user.



Change

Choose from the languages installed on your system or use the Windows

language system default.



DbCapture

About DbCapture

Overview: DbCapture is used to automate the handling of documents received from business partners. It uses Optical Character Recognition (OCR) to read information from incoming documents, validate it and pass the information to an integrated system, such as your financial or ERP package.

Purchasing cycle

A common use of the software is processing incoming purchase invoices from suppliers and the automatic creation of purchase ledger transactions in the integrated finance system.

Sales cycle

Many businesses use the software to process sales documents through the sales cycle from order, delivery, through to invoice and payment.

Capture queue

Usually the incoming documents are added to the DbCapture queue following scanning by use of the [DbCapture export client](#). If documents are received in digital form such as PDF, they can be added directly to the queue without scanning using the optional [PDF module](#).

Once in the DbCapture system the queue may be managed using the [DbCapture admin client](#) and individual documents are processed using the [DbCapture interactive client](#).



The software handles multi-page documents, invoices received as PDFs, foreign currency invoices, foreign language invoices, credit notes, hand-written documents, etc. If there is a purchase order number on the invoice, with many systems we can automate matching and avoid the need for further approval.



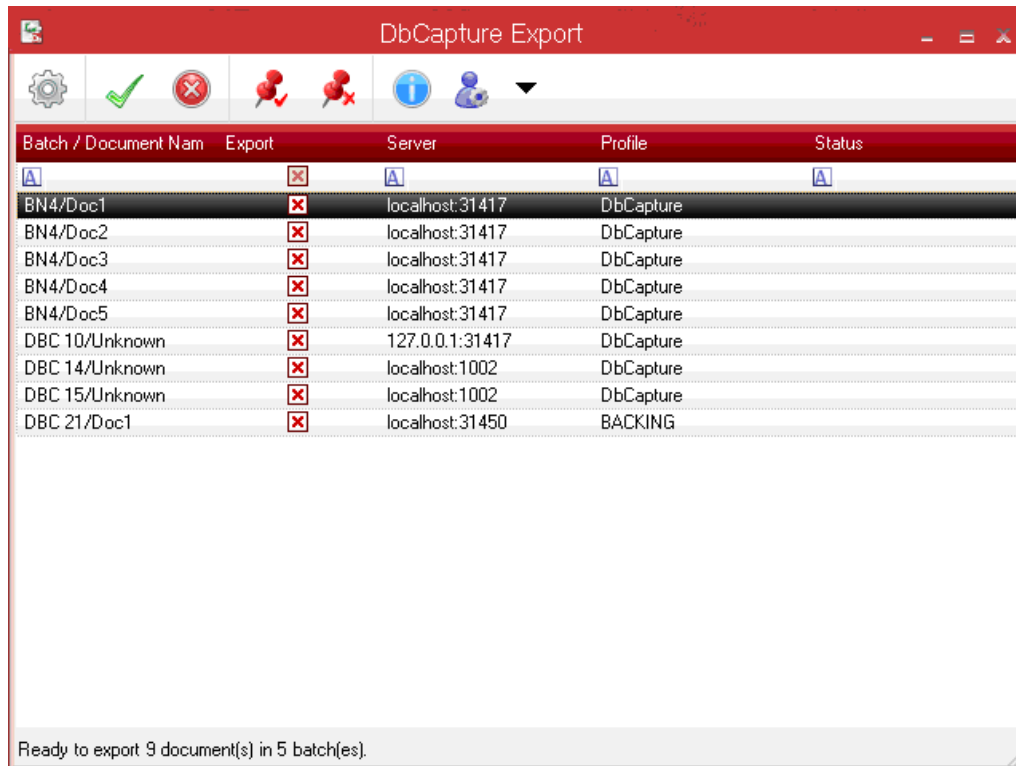
Suppliers can be recognised by a unique VAT number. Account details can then be retrieved from the ledger (the supplier code and supplier name). Key data is extracted from the document using OCR (based on the template from previous invoices from the same supplier). Due date can be calculated from the payment terms in the linked finance system. Data can be validated, including custom rules.


Quality of originals

Many things can affect the quality of the scanning OCR — photocopying, manual faxing, poor print quality, handwritten annotations, rubber stamps on the originals, misaligned original stationery, etc. Some of these can be controlled by your business procedures to improve the accuracy and reliability of data capture, eg don't use a 'Date Received' stamp over the top of the invoice header information.


About DbCapture export

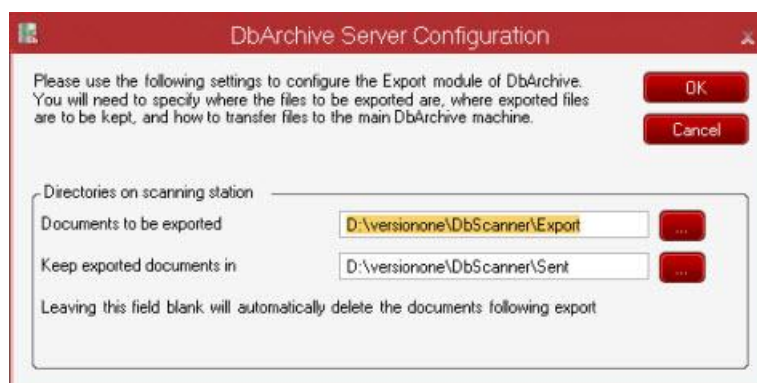
Overview: DbCapture Export is used to export documents scanned using DbScanner into the DbCapture queue. It is launched automatically from DbScanner when the export option is selected.



 DbCapture Export has the same functionality as DbArchive Export.

Button	Description
--------	-------------

	Configure the export server. Choose a folder for your export documents. You can also choose to keep exported documents in a separate folder (like Sent). Leaving this field blank automatically deletes documents after exporting.
---	--





Process — export the document batch.



Cancel the export and close the window.



Select All — select all the documents available for export. You can then clear individual documents from the list. Use a combination of this and the next button to get the selection you want.



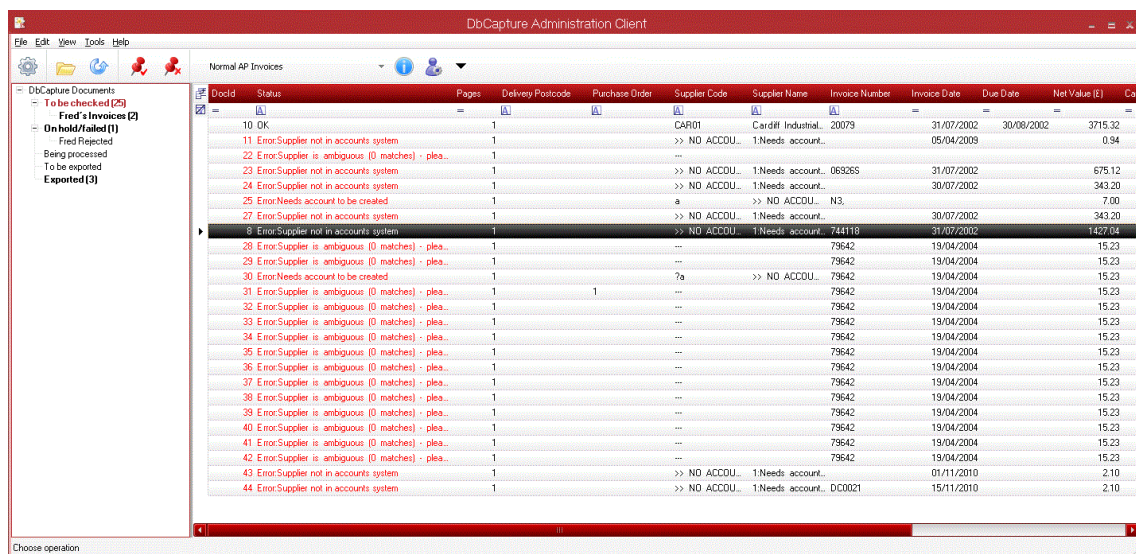
Select None — clear the selection of all documents available for export. You can then select individual documents from the list. Use a combination of this and the previous button to get the selection you want.

About DbCapture admin client

Overview: the DbCapture Admin Client is used to manage the entries in the DbCapture queue.

The main screen shows the queue entries split by their current status. Select an individual entry and open it by double-clicking or choose the open function to launch the DbCapture Interactive Client and process the entry. The administrator is responsible for managing any entries currently 'On Hold/Rejected' — they are not opened automatically in the [DbCapture interactive client](#). Multiple entries may also be selected and processed as a group.

Other options permit **system administrators** to configure DbCapture.



Filtering the display

Use the status list on the left of the screen to filter the results, eg show just Documents On Hold, To Be Exported, etc.



Processing documents

Menu Option	Description
File Exit	Close DbCapture Admin
Edit Select all	Select All — select all the documents available. You can then clear individual documents from the list.
View Show hidden fields	Show all fields not currently displayed (you can display only the fields you want using the Field Chooser).



Refresh Refresh to update the display to show the latest documents.

Tools Server connection Show the server connection details.



Accept Accept the document as processed.

Put on hold Put the document on hold, perhaps for further investigation. You must enter a reason.

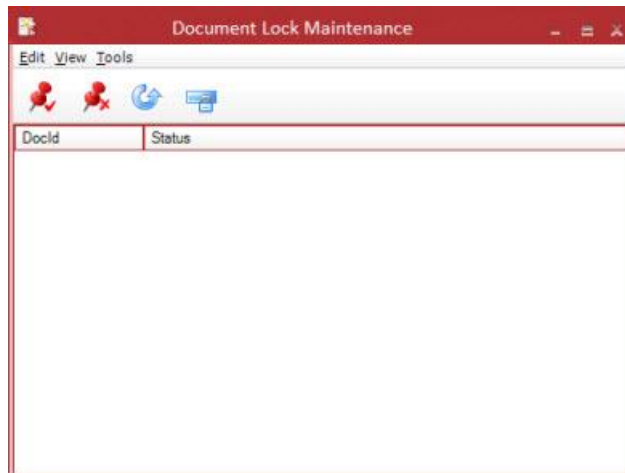
Reinstate Take a document off hold and put it back in the regular queue.

Revalidate Check the document again.


Reprocess Reprocess the document through the capture routines again.

Reprocess failed entries Reprocess just the failed entries.

Document lock maintenance Document lock maintenance is used to release documents left locked when multiple users attempt to access the same document. Occasionally documents can remain locked unnecessarily, use this feature to release them.



Toolbar

Button	Description
	<p>View settings for:</p> <ul style="list-style-type: none"> DbCapture client — the location of the DbCapture Client executable, eg d:\Program Files\Version One\DbCapture Client\DbCapClient.exe Update interval, the interval for polling update. Choose from <ul style="list-style-type: none"> off (manually update only) 15s 30s 1min 10min 30min 1 hour. Document viewing priority: <ul style="list-style-type: none"> default Errors, warnings, OKs Errors, warnings then OKs Errors, warnings only Errors then warnings then OKs Errors then warnings only.



Open the next document and view it in the [DbCapture client](#).



Refresh the display to show the most recent information.



Document filter — choose the document type from the drop-down list.



Select All — select all the documents available. You can then clear individual documents from the list. Use a combination of this and the next button to get the selection you want.



Select None — clear the selection of all documents available. You can then select individual documents from the list. Use a combination of this and the previous button to get the selection you want.

Sub-queues

Your DbScanner queues can have sub-queues, eg within the To be Checked or On Hold/failed lists, you might have sub-levels such as Jack's Documents, Ellie's Documents, etc. The administrator can drag and drop entries between sub-queue lists.

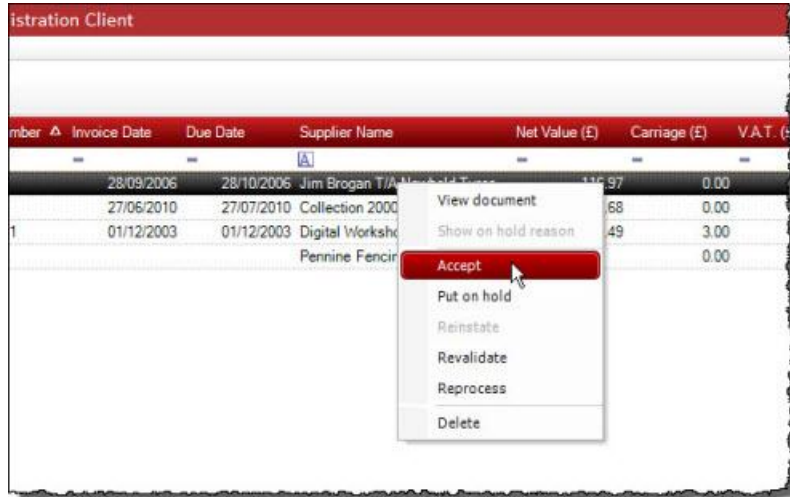


Creation of sub-queues (Jack's Documents, Ellie's Documents, etc) is part of your implementation and not covered here. You can attach permissions to sub-queues, so Jack only sees his own documents and Ellie only sees hers. Administrators would have access to all queues to move documents between them.

There are also options on the [right-click context menu](#) for working with documents.

DbCapture admin context menu

Overview: you can process documents in DbCapture Admin using the mouse right-click options. Some options are unavailable, depending what stage the document is at.

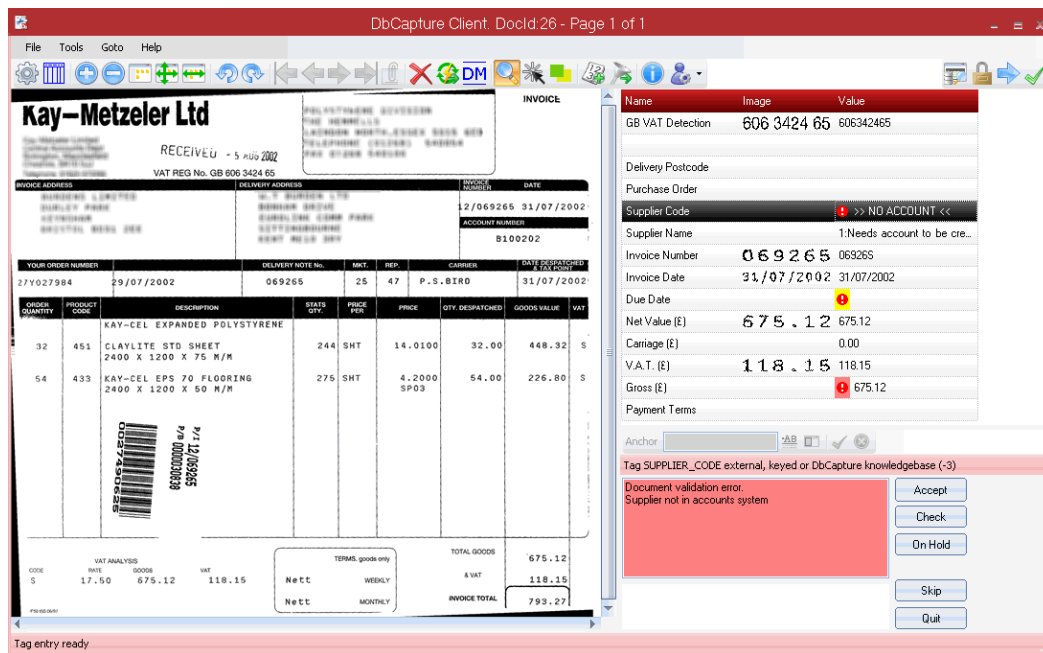


Option	Description
View document	Look at the document in the DbCapture client .
Show on hold reason	Show the reason the document has been put on hold.
Accept	Accept the document as processed.
Put on hold	Put the document on hold, perhaps for further investigation. You must enter a reason.
Reinstate	Take a document off hold and put it back in the regular queue.
Revalidate	Check the document again.
Reprocess	Reprocess the document through the capture routines again.
Delete	Delete the document. You must enter a reason.

About DbCapture interactive client

Overview: the DbCapture Interactive Client is used to process individual entries from the queue. It can run in two modes, if launched from the DbCapture Admin Client you process only the selected entry and it then terminates. If run directly it presents each queue entry in turn until the queue is empty.

From the client screen you can, subject to access permissions, create or amend existing [document templates](#), correct and validate data that has been read from the document and accept the queue entry such that it updates the integrated system.



Menu	Description
------	-------------

File	Displays the File menu, with a single Exit option to close DbCapture.
------	---

Tools	Display the Tools menu:
-------	-------------------------

Option	Description
--------	-------------

Server connection	Server connection details.
Delete document	Delete the displayed document.
Save user data	Save user data.
Secondary identification	Secondary identification. Use to add a secondary identifier for the document. For example if you use VAT as the primary identifier in a multi-branch or multi-company environment, you might need to distinguish between documents from a particular branch address. Ideally the code is be unique, so you might use the










	phone number or possibly the postcode.												
Display anchors	Display anchors. Show the positions of tagged data fields.												
Extract tags	Extract tags.												
Extract line data	Extract line data. If enabled, you can extract further line level detail from your documents. Line level capture is an extra option, enabled during setup and also switched on or off at template level. Line level information is shown at the bottom of the screen.												
Filter document	Filter document.												
Extraction processing	Extraction processing, shows a sub-menu with further option.												
	<table border="1"> <thead> <tr> <th>Option</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Reset tags</td> <td>Reset your tags.</td> </tr> <tr> <td>Reset line level extraction</td> <td>Reset any line level extractions for this document (if enabled).</td> </tr> <tr> <td>Reset all extractions</td> <td>Reset all extractions at both document and line level (if present).</td> </tr> <tr> <td>Re-identify and extract</td> <td>Re-identify the document and extract the data again.</td> </tr> <tr> <td>Undo all changes</td> <td>Undo all changes to the document in the current session.</td> </tr> </tbody> </table>	Option	Description	Reset tags	Reset your tags.	Reset line level extraction	Reset any line level extractions for this document (if enabled).	Reset all extractions	Reset all extractions at both document and line level (if present).	Re-identify and extract	Re-identify the document and extract the data again.	Undo all changes	Undo all changes to the document in the current session.
Option	Description												
Reset tags	Reset your tags.												
Reset line level extraction	Reset any line level extractions for this document (if enabled).												
Reset all extractions	Reset all extractions at both document and line level (if present).												
Re-identify and extract	Re-identify the document and extract the data again.												
Undo all changes	Undo all changes to the document in the current session.												
Properties	Show more details for the selected document.												
Goto	Display the Goto menu:												
	<table border="1"> <thead> <tr> <th>Option</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>[Name]'s invoices</td> <td>Display only the named person's invoices. This is a useful way to filter the data with multiple users.</td> </tr> <tr> <td>To be checked</td> <td>Display only documents to be checked.</td> </tr> </tbody> </table>	Option	Description	[Name]'s invoices	Display only the named person's invoices. This is a useful way to filter the data with multiple users.	To be checked	Display only documents to be checked.						
Option	Description												
[Name]'s invoices	Display only the named person's invoices. This is a useful way to filter the data with multiple users.												
To be checked	Display only documents to be checked.												
Help	Displays the About DbCapture Client help menu .												















Capture problems

Any problems with the capture are highlighted in the document, for you to correct. The example screen shows a handwritten invoice that can't be read correctly, the software highlights the field and you can correct it manually.



There is validation logic attached to some of the fields, eg invoice Gross Value must equal the total of Invoice Nett plus VAT.

Icon	Description
	Display the properties for the current document.
	Line level display .
	Zoom in and make the document bigger.
	Zoom out and make the document smaller.
	Actual size. Show the document actual size, use the scroll bars to pan around.
	Full page. Show the document as large as possible whilst still showing the whole document in the window.
	Page width. Fit the document to the page width in the available window, use the vertical scroll bar to pan up and down.
	Rotate the image left a quarter turn (anti-clockwise).
	Rotate the image right a quarter turn (clockwise).

	First page of a list of documents.
	Previous page of a list of documents.
	Next page of a list of documents.
	Last page of a list of documents.
	Display attachments. Show any linked attachments to the current document. DbArchive can hold other types of documents that you haven't necessarily scanned in, eg terms and conditions documents, contract attachments.
	Delete the current document.
	Reprocess the document. If you're not happy with the results from the first pass, you could try re-scanning for better results.
	Reprocess with dot matrix bias. Reprocess the document to try and improve the results, using pre-sets optimised for documents originally printed on dot matrix printers.
	Zoom mode. This lets you zoom right in a particular area of a document. The cursor changes to a magnifying glass, use it draw a box around the area you want to look at. Use the zoom out button (above) to shrink the display again.
	Identify fields.
	Display zones. Use to show the areas of the document recognised as holding field data. Data that has been recognised and looks ok is shown with a green background. Data with a warning is shown with a yellow background — these fields need checking and possibly amending. Data with a more severe problem is shown with a red background — these fields definitely need correcting.
	Enable secondary identification. Use to add a secondary identifier for the document. For example using VAT as the primary identifier, you might want to distinguish between documents from a particular branch address. Ideally the code would be unique, so you could use the phone number or possibly the postcode.
	Filter attachments.
	Display the help about window.

Validating data

Check the data you are validating by clicking on the field in the main body of the window. The status is shown in the box at the foot of the window.

Button Description

Accept	Accept the document and save it.
Check	Check the document details.
On hold	Put the document on hold without passing it.
Skip	Skip this document and display the next in any queue.
Quit	Exit without saving.

The same functions can also be called from the toolbar at the top right of the window.



DbCapture handles all documents, including those hand-written. Handwriting is not OCR'd automatically. DbCapture also gives unreliable results with decorative display fonts.

See also:

[Capture templates](#)

[Understanding document identifiers](#)

[Creating and using anchors](#)

Capture templates

Overview: when you scan a document, the software uses intelligent recognition to identify the relevant fields of the document and work out the values in the fields.

For example, the software can recognise the gross value, because it is a numeric value, normally found in the same position at the bottom right of a document. DbCapture uses this information to make the recognition job easier by predicting where data is likely to appear on a document and what the data value is.

Information identified using the template can be used to lookup up unique detail about a customer or supplier in a linked application, eg your financial package or ERP. So you might be able to use the VAT number to identify a supplier, or a phone number to identify a particular customer branch.



System administrators only: we do not recommend editing templates unless you are competent and sure of the implications. Take advice if necessary.

Installing templates

Templates are normally installed for you when the software is first installed and commissioned.

Editing templates

Contact Version One if you want to edit a capture template. Editing the capture templates with out a complete understanding risks the integrity of your system.

See also:

[About DbCapture interactive client](#)

[Understanding document identifiers](#)

[Creating and using anchors](#)

Understanding document identifiers

What is a document identifier?

Document identifiers are used by DbCapture to associate a scanned image with the template, to tell it where to find the information in a document. The identifier is a piece of information which can be read from the document by DbCapture.

The most common identifier for purchase invoices is the supplier's VAT number. If this isn't present on the document or can't be read successfully, use another identifier.

You have two identifiers for a document, to handle special cases (see below). There's always a Primary Document Identifier, appearing in the top row of fields in the DbCapture client.

Name	Image	Value
GB VAT Detection		598323109
Delivery Postcode		
Purchase Order		

If necessary you can also manually create a Secondary Document Identifier, appearing on the second row of the fields, post code in this case.

Name	Image	Value
GB VAT Detection		598323109
Secondary Identifier		S43 2PR
Delivery Postcode		
Purchase Order		
Supplier Code		NO ACCOUNT

Using document identifiers

When a document is scanned and exported to DbCapture the information on the scanned image is read using OCR (Optical Character Recognition).

DbCapture searches for the most likely document identifier, using built-in rules. When it finds the most likely, it checks whether it already has a template stored for that document. If so, it loads the template and extracts information from the document to show on the [DbCapture client](#).

If there's no template it can't extract any more details and you need to create a template for that document.

Creating templates

This involves showing DbCapture where on the page it should look for particular pieces of information, eg invoice number, invoice date, etc.

Once DbCapture knows where to look for information, this knowledge is saved as a template linked to that document identifier value, eg the supplier VAT number.



Field location is relative to the document identifier, eg invoice number might be 2 cm to the left and 3.5 cm below the identifier.

Special cases

Sometimes DbCapture may not find a suitable document identifier or it may not uniquely identify the template. You then need an additional way to create identifiers.

No automatic identification

If DbCapture does not find a suitable candidate for document identification you need to do it manually using the client. Good identifiers:

1. Unique to a particular document and supplier, ie not something common like 'Invoice'
2. In a fixed location on the page, ie the same place on every document
3. Unlikely to be misread, ie short and easy to read, ideally with white space around
4. Not a picture or a company logo using a decorative font.



The identifier needs to be reliable so all documents from that supplier are identified correctly.

Common manual identification fields are supplier postcode, fax or telephone number. Other fields may also be suitable, look at several documents from the supplier before deciding.

[Show/hide example.](#)



If you manually select the identifier, DbCapture will not automatically be able to identify the supplier code and name at the point at which you create the template. You need to use the Query Field to lookup the suppliers and select the correct one. Once you've done this and stored it in the template it is known to DbCapture in the future.

Primary document identifier not unique

Sometimes the identifier found by DbCapture does not uniquely identify the template and supplier, eg:

- suppliers with a single VAT number for multiple trading names and document formats
- suppliers with multiple account codes in your purchase ledger, where you need to identify which code the invoice should be posted to.

To deal with this you need to 'disambiguate' — select and create a secondary document identifier.

To disambiguate a document you need to find something which can reliably separate it from other documents sharing the same primary identifier. The rules for doing this are the same as those for a manual primary document identifier. If necessary, click the Disambiguate button on the toolbar, which prompts you to select the secondary identifier. If this has been done for you, you don't need to click the button.



Once you have created the secondary document identifier and confirmed the supplier code to which the document relates, this is stored with the template and allows documents to be recognised automatically in the future.



When a secondary document identifier has been created on a template, the location on the document where DbCapture should look for each field is stored as a position relative to this secondary document identifier rather than the primary identifier.



Secondary document identifiers should **only be created when absolutely necessary**, ie when the primary document identifier does not uniquely identify the template and the supplier.

Summary

Document identifiers tell DbCapture where to find the information on a document using a series of templates. Careful creation of templates and the selection of document identifiers is key to successfully processing documents.

See also:

[About DbCapture interactive client](#)

[Capture templates](#)

[Creating and using anchors](#)

Creating and using anchors

What is an anchor?

Apart from the primary document identifier (usually the VAT number) each field extracted from the document using OCR has an anchor. The anchor provides DbCapture with a fixed point from which to navigate to the data field to be extracted.

By default all extract fields are anchored to the primary document identifier. If there is a secondary document identifier, this becomes the default anchor for all extract fields. You can see the anchor for any field by turning on Display Anchor Mode (choose Tools > Display Anchor), and positioning the cursor in the extract field box. The extract area for the field is shown in Green and the Anchor in Yellow.

What is a floating anchor?

Floating anchors are used when text to be extracted moves from invoice to invoice, eg line values on an invoice may move up and down depending on how many lines there are.

They may also be required when the default anchor is not suitable for other reasons, eg the VAT number is pre-printed onto continuous stationery and the paper is not always correctly aligned.

In these situations the extract fields must be anchored to some text whose position remains constant in relation to the extract field.

What makes a good floating anchor?

Floating anchors are always searched for within a marked region. Good floating anchors:

1. Must not float in relation to the extract field anchored to them
2. Unique within the marked region, eg don't use 'Total' if both 'Total Goods' and 'Total VAT' appear in the marked region
3. Short and likely to OCR correctly, eg avoid multi-word phrases, punctuation characters, etc.

If multiple fields require a floating anchor, try and use the same one for all fields.



If a floating anchor is not found when a document goes through the OCR, all fields using it are blank and the green box appears top left of the image.

Creating floating anchors

Steps to extract the Gross Value and create an anchor that locates the gross value field:

1. Select Gross value and extract it as normal.

	=====		Ancillary Charge 6 (-)	0.00
Net	3209.33		VAT Value	0.00
ce @ 17.50%	561.63		Gross Value	3770.96
TOTAL	3770.96		Calculated Net Value	0.00

2. Right-click the value column of the gross value, choose Display Anchor Mode.

Gross Value	3770.96	3770.96
Calculated Net Value		0.00
Calculated Variable Value		0.00
VAT Code		NVA
VAT Rate		0.0

3. Right-click, choose Attach Anchor To Field, then Create New Anchor.

VAT Value		0.00
Gross Value	3770.96	
Calculated Net Value		
Calculate		
VAT Code		

4. The right of the window greys-out, use the box bottom left to create the anchor.

Select anchor text ← → ✓ ✕

5. Grab the text from the screen you wish to use as your anchor. This is a bit of text which follows the values up and down the page, in this example the word 'Total' follows the text up and down.

```

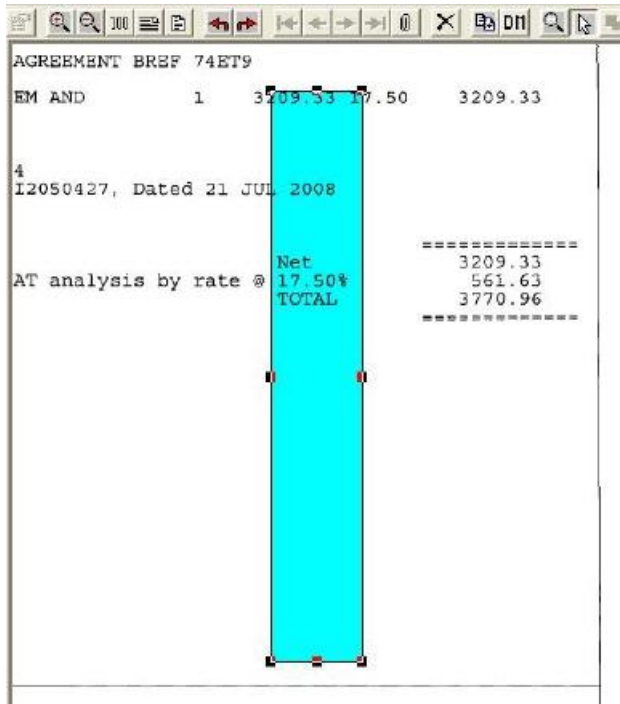
=====
Net                3209.33
by rate @ 17.50%  561.63
TOTAL              3770.96
=====
    
```

6. Extract the text as normal, it is placed in the bottom box.

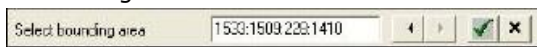
Select anchor text ← → ✓ ✕

7. Click the right arrow.

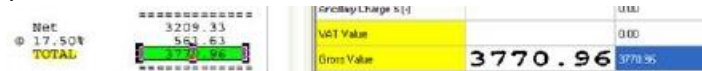
- Tell DbCapture the boundary for the search word. This is an area around the word 'Total' which extends up and down the page — draw a box on the page (the co-ordinates are displayed in the box).



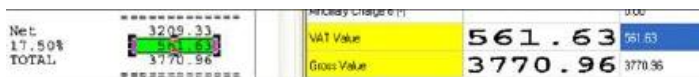
- Click the green tick.



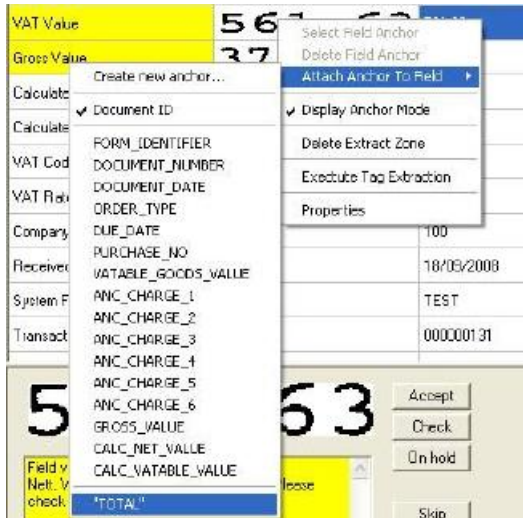
- Now the Gross Value row is selected, the extraction zone is green and the anchor text yellow.



- To attach the rest of the values to that anchor point, select the VAT Value and extract it as normal.



12. Right-click the Value column of the VAT Value row and select Attach Anchor To Field to see your 'Total' anchor. Select Attach To The Anchor.



13. The floating anchor has been created.

Net	3209.33	Ancillary Charge 6 (%)	0.00
17.50%	561.63	VAT Value	561.63
TOTAL	3770.96	Gross Value	3770.96
		Calculation Net Value	0.00

See also:

- [About DbCapture interactive client](#)
- [Capture templates](#)
- [Understanding document identifiers](#)

Line level capture

Overview: line level capture is an optional feature within DbCapture, used to gather a greater level of detail from the documents using either using OCR or data obtained from elsewhere in your Finance or ERP system.



Most organisations and users **don't use line level capture** and can skip this topic.

Line level capture is enabled in two ways:

- switched on globally at the time of implementation licensing
- if it is enabled globally, it can also be switched off individually for each template.

If it is not switched on, you won't see the line level options.

Line level grid

Line level capture uses an additional line level data grid, below the document image in the window. Columns in the grid are configured in the DbCapture profile.

The grid data can be populated in three ways:

1. Manually — type in all data
2. OCR — the default contents are read from the source document using OCR. This requires you to define the extraction zones and rules in the document template.
3. From your other computer systems — configured as part of the initial project setup.

However it is created, data in the grid may be amended, lines added and deleted and the entire contents validated.

See also: [Using line level capture](#)

Using line level capture



Most organisations and users **don't use line level capture** and can skip this topic.

Once you've identified document header details, you can run line level capture. Line level capture is used to gather a greater level of data about documents.



Line level capture has its own toolbar towards the bottom of the window. If the toolbar and data grid are not visible you may need to click the View Line Level Grid icon on the main toolbar. The icon is used to view and hide line level details at any time.

Line level toolbar:



If the data grid is populated via OCR, you may need to create the extraction details if they are not already present. To do this:

First page

1. On the bottom tool bar click the Edit Line Data icon.
2. On the first page containing details, click and drag a box over all the columns — this is the boundary box.
3. Right-click in the boundary box and choose Properties.

Payment has been made for Housing Benefit for the following tenant(s) by Bank BACS on June 8, 2011

Posting id: 883295 Amount: £2,788.55 Period: 16-MAY-2011 / 12-JUN-2011

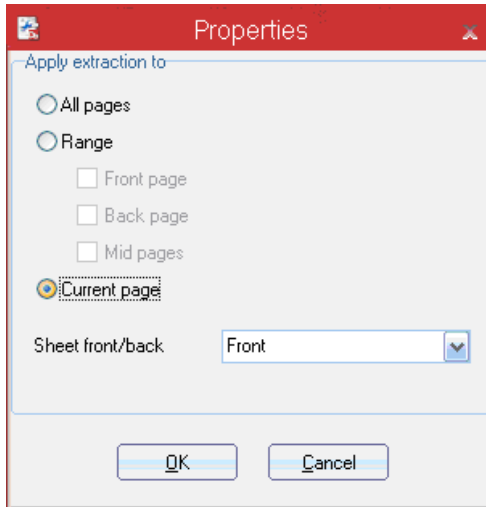
Claimant Details
Address: *Prapsif*

Claim ID	Claimant	Address	Amount	Start	End
176343	D PISANO	2, MADDOCK LIDGE	£350.31	16-MAY-2011	12-JUN-2011
176603	D WHYTE	11, BRIDGEMAN LANE	£335.24	16-MAY-2011	12-JUN-2011
183373	J KING	11, BRIDGEMAN LANE	£384.16	16-MAY-2011	12-JUN-2011
203153	C HAMPSON	11, BRIDGEMAN LANE	£495.80	16-MAY-2011	12-JUN-2011
208984	F LUSHER	11, BRIDGEMAN LANE			
211103	M PETTITT	11, BRIDGEMAN LANE			
214236	B FRYATT	11, BRIDGEMAN LANE			

Context Menu:

- Edit columns
- Select top anchor
- Attach top anchor to table
- Delete top anchor
- Select bottom anchor
- Attach bottom anchor to table
- Delete bottom anchor
- Display anchor mode
- Delete
- Properties

- If the first page of details has them in a different place to all the other pages, select current page, use the range option or leave it on all pages.



- The Sheet Front/Back drop-down list is for duplex — if the first side of the first page is different, choose Front. Otherwise, select front & back to process both the same.
- Click OK to continue

Next page

- Go to the next page, and drag a boundary box over the line details.
- Right-click in this boundary box, and set the Properties. If it's single sided and all the continuation pages are the same.

Define columns

Now define the columns:

- Go back to the first page.
- Right-click on the boundary box, and choose Edit columns.
- Click on a column you use to identify the line items, eg Claim Ref, Line Item, or whatever data your documents hold. This highlights the box in the Add New row with a dashed line.
- Drag a box over the column on the page containing the line identifier (eg Claim Ref, Line Item, etc).
- Choose extract lines or repeat the above for the remaining extractable columns. This extracts the text on each line from the document.
- Select the column you use to identify the line items, then drag click and drag a box for the next column you want to capture, eg description, value. Repeat for each column you want to capture.

Extract line details

Click Extract lines to add the new columns into the grid.

Viewing columns

To see all the columns you have extracted on the page, right-click the current column in the image viewer and select 'View all columns'. This shows the current column in grey, other columns in blue.

Landlord Payment Schedule

LONDON BOROUGH OF RICHMOND UPON THAMES

Civic Centre, 44 York Street, Twickenham TW9 3JZ
 Website: www.richmond.gov.uk

Telephone: 0208 891 7724
 Date of issue: 08-Jun-2011

email: revenue@richmond.gov.uk

If you need to contact us please quote this reference: **216**

HANOVER HOUSING ASSOCIATION
 Creditor ID:
 Payment Method: Bank BACS

65,499.85 //

Dear Sir / Madam,

Payment has been made for Housing Benefit for the following tenant(s) by Bank BACS on June 8, 2011

Posting lit: 883295 Amount: £2,788.55 Period: 16-MAY-2011 / 12-JUN-2011

Claimant Details
Prop 95r

Claim ID	Claimant	Address	Amount	Start	End
176343	D PITCHER	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£250.31	16-MAY-2011	12-JUN-2011
178663	D WHITE	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£335.24	16-MAY-2011	12-JUN-2011
183373	J KING	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£304.16	16-MAY-2011	12-JUN-2011
203153	C HARRISON	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£495.80	16-MAY-2011	12-JUN-2011
208694	J UGHELLI	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£391.72	16-MAY-2011	12-JUN-2011
211103	M PETERSON	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£433.96	16-MAY-2011	12-JUN-2011
214236	B FRISVOLD	10, BRIMLEY COURT, TWICKENHAM, MIDDLESEX, TW9 3JZ	£487.36	16-MAY-2011	12-JUN-2011

HB Ref	Tenancy ref	Tenants Surname	HL
176343			
178663			
183373			
203153			
208694			
211103			

Edit grid bounding

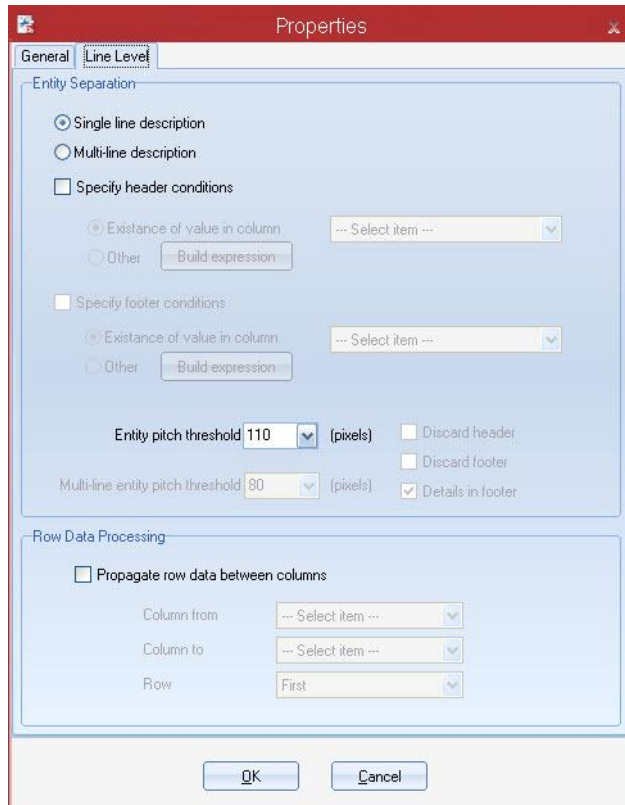
View all columns

Delete

Troubleshooting

Adjusting the line pitch

1. If the extraction only returns a single row, try altering the amount of space between the lines. From the Tools menu, select Properties and choose the Line Level tab.



2. Choose a higher Entity Pitch Threshold using the drop-down list (for 300dpi scans each linear inch is 300 pixels).
3. Click OK for the Properties window and click Extract Lines again.

Deleting columns

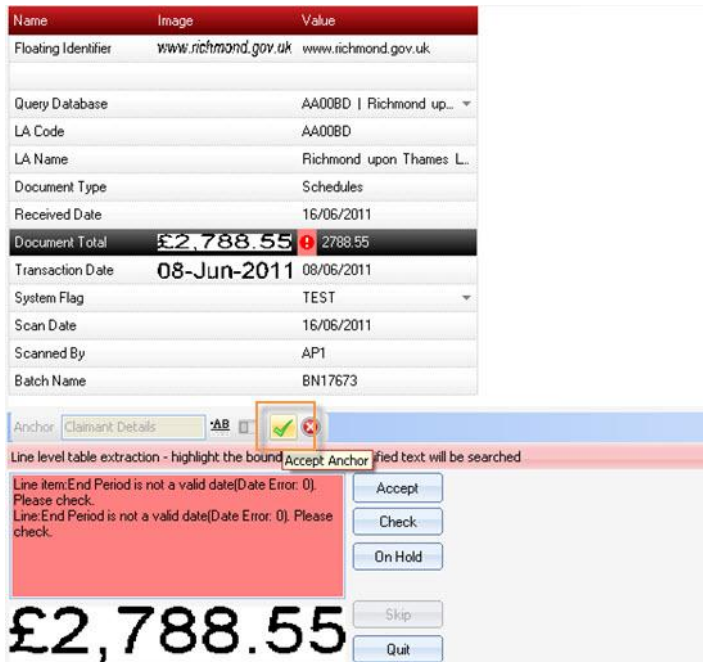
To delete extra grid lines manually, select the line, then click the red cross in the bottom toolbar. Use Extract Lines to re-extract this data if you make a mistake.

Adjusting the grid boundary

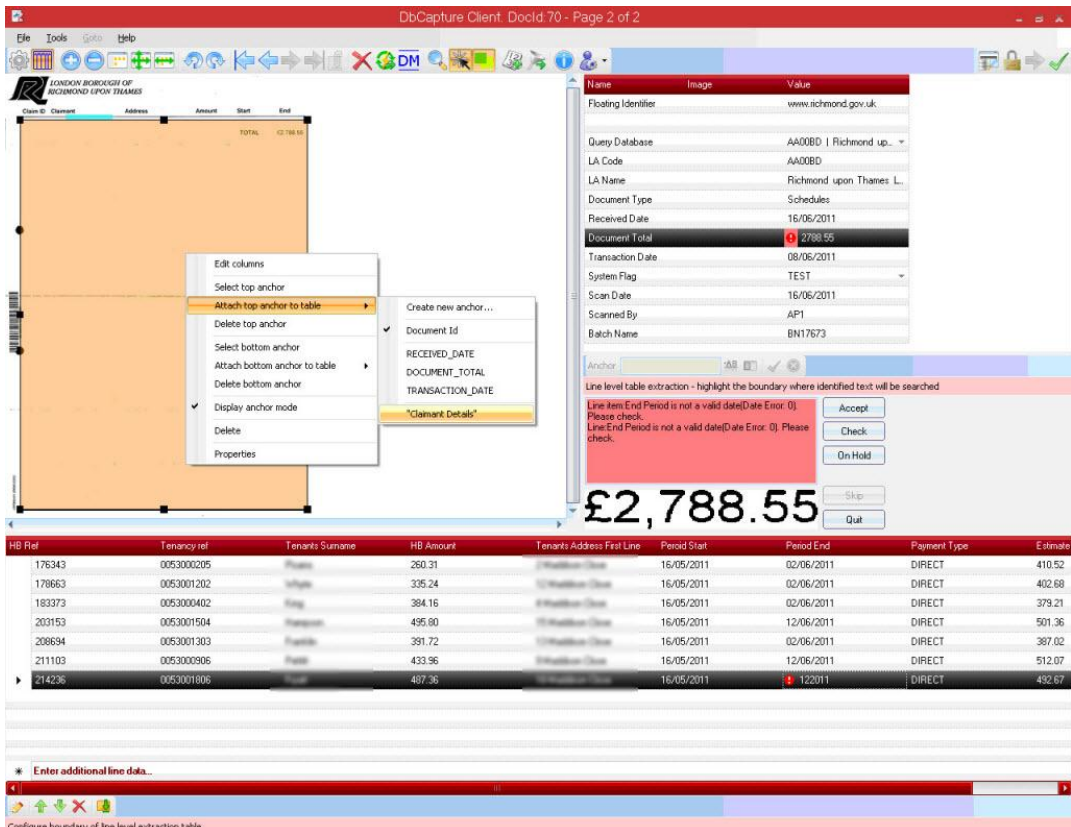
To switch back to the boundary editing mode, right-click the current column and choose Edit Grid Bounding.

Adding start and end markers

1. In boundary editing mode, right-click the boundary box and make sure Display Anchor Mode is selected.
2. Right-click the boundary box and choose Attach Top (or Bottom) Anchor To Table, then Create New Anchor...
3. The Anchor toolbar opens. Select the text to use, eg line details. Click and drag a box over the text.
4. Configure the boundary to look for this text, using the Configure boundary button on the Anchor toolbar. Click and drag a box on the page where this label text might appear.
5. Click the green tick to accept the anchor.



- Right-click the boundary box and from the appropriate Attach To Table menu, choose your new anchor from the bottom section of the list.



- Repeat for each boundary box you have configured, eg the second page might have a total you could use to define the bottom of the boundary box.

Creating a new anchor

Select the text, click Configure Boundary, click and drag a box over the column where it might appear. You can adjust the boundary box by dragging the handles if necessary.

Shared anchors

If any boundary box shares an anchor, ensure your boundary box encloses both, and select them from the Attach Anchor menu rather than trying to create them again. If the line details have additional lines between them, you can configure this under the Tools > Properties menu.

Extracting data manually

Once the template extraction zones have been defined in the template, or for systems where OCR is not being used to put content into the data grid, the data in the grid can be manipulated manually.

Data can be entered or amended in each of the cells on a specific detail line simply by clicking in the cell and typing as normal.

Deleting a line



You can delete a line from the grid by selecting it and clicking the Delete icon (red cross) on the toolbar.

Adding a line to the grid

To add a line to the grid, put the cursor in the Enter Additional Line Data section and edit the required cells.



Moving lines



To move lines up and down within the data grid, select the line you want and click the Up or Down arrows on the toolbar.

Extracting lines



At any point, if the grid data has been populated via OCR you can completely refresh the grid and re-extract the data from the document by clicking the Extract Lines icon on the toolbar. This is useful when creating the template, enabling you to tune the individual line extraction rules.



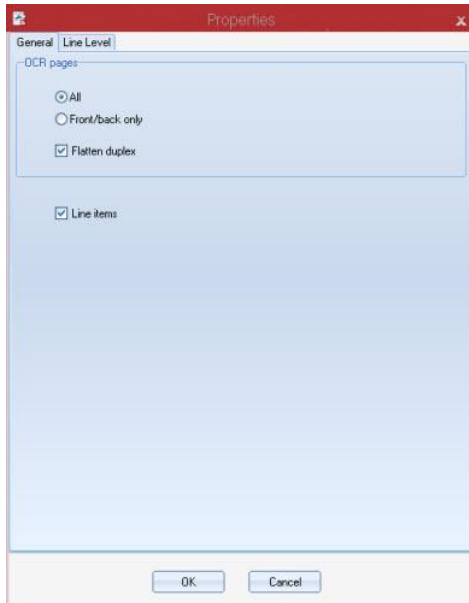
Any manual changes you made are lost when you extract lines.

Capture properties

Overview: properties of a DbCapture document are shown by clicking the Properties toolbar button from the [DbCapture interactive client](#) window. The properties window has two tabs:

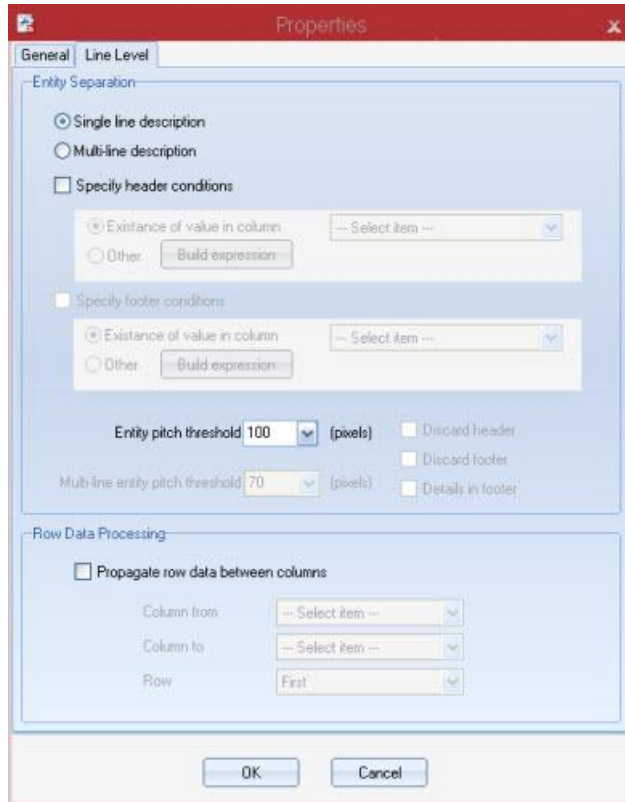
- general
- line level.

General tab



Option	Description
All	Choose to process all pages of a document. Select this or the next option.
Front/back only	Choose to process just the front and back pages. Select this or the previous option.
Flatten duplex	
Line items	Select whether you want to process line level details for this document.

Line level tab



Option	Description
--------	-------------

Entity separation

Single line description	Choose to specify whether line descriptions are typically single line.
Multi-line description	Choose to specify whether line descriptions are typically multi-line (not necessarily all descriptions).
Specify header conditions	Select to choose header conditions. You then need to select one of the two options below.
Existence of value in column	<p>Choose from the drop-down list:</p> <ul style="list-style-type: none"> • number • description • quantity • unit • price • discount % • value. <p>Alternatively choose Other (below).</p>

Other	If you choose Other (rather than Existence of Value in Column above), press the Build button to specify the condition.
Specify footer conditions	Select to choose footer conditions. You then need to select one of the two options below.
Existence of value in column	<p>Choose from the drop-down list:</p> <ul style="list-style-type: none"> • number • description • quantity • unit • price • discount % • value. <p>Alternatively choose Other (below).</p>
Other	If you choose Other (rather than Existence of Value in Column above), press the Build button to specify the condition.
Entity pitch threshold	Choose off or a value from the drop-down list (between 100 and 300 pixels).
Multi-line pitch threshold	Choose off or a value from the drop-down list (between 50 and 100 pixels). Only enabled if Multi-Line Description is selected at the top.
Discard header	Select to discard the header information from the document.
Discard footer	Select to discard the footer information from the document.
Details in footer	Select to specifically look for details in the footer of the document.

Row data processing

Propagate row data between columns	Select and choose the data you want to propagate from the fields below.
Column from	<p>Choose from:</p> <ul style="list-style-type: none"> • number • description • quantity • unit • price • discount % • value.
Column to	Choose from:

-
- number
 - description
 - quantity
 - unit
 - price
 - discount %
 - value.

Row Choose first or last from the drop-down list.



The line level tab is only available if [line level capture](#) is enabled.

PDF import

PDF import module

Overview: this is an add-on module to DbCapture that can be used to import files that already exist having been created elsewhere. These may have been received via email from your business partners. The most common format of these files is PDF, other formats such as TIFF and JPEG are also supported. The files are automatically imported and processed in DbCapture, without needing DbScanner.

Technical outline

The default TAG values such as LIVE/TEST or COMPANY NUMBER, etc, which are normally passed from DbScanner can be automatically populated based on settings in a file in the polled directory. The TAG fields can be generated and specified based on the network folder from which the PDF or TIFF file is obtained. For example:

`c:\dbcapture\pdf\test` generates documents to go into the test system

`c:\dbcapture\pdf\live` generates documents to go into the live system

You can monitor a maximum of ten folders. If you need more tags, they can be manually entered with the DbCapture client.



This module requires a minimum of DbCapture v2.1 and ABBYY 9. We recommend upgrading all the DbCapture suite to the current version when implementing this module. The module is separately licensed and requires a valid 'AutoDeposit' entry in the DbCapture.key file. The previous key entry 'PDF Support' remains valid for backwards compatibility.

Using the PDF module

How it works

Documents which have been received in digital format are placed by the operator into a specified folder on disk. Multiple folders may be configured as a configuration file present in the folder can be used to sets specific options in the entry that will be added to the DbCapture queue. The module periodically checks folders) for input files. These are converted into an image format which can be processed. Files are put in the DbCapture queue [To Be Processed] as if they had been scanned and exported.

The invoice is then passed through the same process as a scanned invoice:

1. The document is OCR'd
2. The data is validated
3. It is put in a DbCapture queue (with an OK, Warning or Error status)
4. If OK, it can be accepted
5. If an error or warning is shown, check the document as normal
6. Data is exported to the finance system
7. The original file is placed in DbArchive (to comply with audit, as this was the original document record).

Installing and configuring

Installation

Initial installation generally uses the IManager utility. After installing the base DbCapture elements in the normal way, select the PDF Module tab and follow the instructions in the installation guide.

Configuration

Installation Manager (IManager) has been updated to allow configuration of polled directories and tag defaults.

The default TAG values such as LIVE/TEST or COMPANY NUMBER, etc, which are generally passed from DbScanner can be automatically populated based on settings in a file in the polled directory. The TAG fields can be generated and specified based on the network folder from which the PDF file is obtained. For example:

`c:\dbcapture\pdf\test` generates documents to go into the test system

`c:\dbcapture\pdf\live` generates documents to go into the live system.

You can monitor a maximum of ten folders. If you need more tags, they can be manually entered with the DbCapture client.

DbCapturePDFFileWatcher

This new service utility polls the configured folders and launches a batch file to add documents it finds to the DbCapture queue. This is an enhanced version of DbCheck and works in a similar manner.

DbCapUtil command line utility

This utility can be used to take action on DbCapture items from the Command Line. For the purposes of this module a new option 'DEPOSIT' has been implemented.

- Install using DbCapUtil.msi, the utility can be executed from any location however it needs the file DbCapUtil.exe.config to exist in the sme folder
- Run `DbCapUtil -h` to display the command line options (see -h snippet below).

```
>>DbCapUtil version 2.1.0.0
```

```
-s          DbCapture server name
```

```
-o          DbCapture server port
```

```
-u          DbLogin user name
```

```
-p          DbLogin user password
```

```
-dbcp      DbCapture profile
```

```
ACCEPT|DELETE|DEPOSIT|QUERY
```

- dbcs DbCapture scanner profile]
- y Document type (PDF, TIF, JPG OR DL)
- doc document to deposit file name
- t tag data file name
- p document unlocking password (eg PDF password protection)
- d DbCapture system document ID
- message reason for operation, currently only applies to deletion]

Deposit example:

```
DbCapUtil.exe DEPOSIT -s dbcap_svr -o 31450 -u user1 -p pw123 -y TIF -doc
.\inv2.pdf -t .\inv2.tag -dbcp DbCapture -dbcs dbcapture1
```



The following file types are supported: PDF, TIFF, JPEG and DL.

Document background processing

The background processor can be executed to run in 'unpack' mode, where a deposited PDF or TIFF is unpacked into media (.dl + images) that the DbCapture client can process. This is achieved using a similar command line to the snippet below (included within the service installation described earlier).

```
dbcscaptureserver.exe -install "DbCapture:Unpack" -UNPACK -
SERVERLOGd:\versionone\Dbcapture\DbCapUnpack.lg
```

Configuration has been added to the DbCapture scan profile container (.dbcs) that configures whether the OCR is performed on the original deposit source or the unpacked media, called `OCRFromMaster`. There is an additional property that forces a format for the unpacked media called `UnpackedFormat`.

OCRFromMaster (.dbcs [main] section)

- 0 Perform OCR on unpacked media. (Default)
- 1 Perform OCR on original deposit source eg PDF file.
- 2 All pages of source document are converted to colour JPEG.

This configuration currently applies to PDF document types only.

UnpackedFormat (.dbcs [main] section)

- 0 If conversion detects that the underlying page of the source document (eg PDF) is colour then the equivalent unpacked page image format is determined by the conversion process. A colour page detection results in a colour JPEG image conversion, a B&W page detection results in a B&W TIFF group 4 image conversion. (Default).

- 1 All pages of source document are converted to B&W TIFF group 4.
- 2 All pages of source document are converted to colour JPEG.



The default operation is to OCR from unpacked media of format of the most applicable output (decided by the conversion process). This can result in a mixture black and white tiffs and colour JPEGs, so if the unpacked media is to be deposited into DbArchive then a mixture of image format types would be present in a dl container. To overcome this you may want to enforce a format (eg black and white) and also perform OCR on the original source as conversion may cause loss of data.

Passing document type to DbCapture scripts

Properties have been added to DbCapture scripts input spool that convey the document type and the file name of the source-unpacked file called Document_Type and Document_Name. Currently the Document_Type can be either PDF or DL as with the case of scanned documents that have been added to the system queue using DbCaptureExport. (see .spl system page below)

```

TRANSFORM:VALIDATE
ACTION:CHECK
USER:ian.fildes
CURRENT_DATE:23/09/2010
CURRENT_TIME:09:05:54
DBC_STATE:1
TAG:C:\versionone\DbCapture\Queue1\168\168val.TAG
DocID:168
Scanner_Profile:dbcapture1
DbCapture_Profile:dbcapture
Template_Type:0
Template_ID:0
Template_Subtype:0
Document_Type:PDF
Document_Name:C:\versionone\DbCapture\Queue1\168\168.PDF
LineCount:0
Template_Userdata:
UserData_Primary:0

```

PDF importing FAQs

Some frequently asked questions about the PDF import module:

Q: What format of PDF files are supported?

A: PDF files can be of either embedded text or pure image format.

Q: What format of TIFF files are supported?

A: The rules are the same as previously applied to TIFF files imported via DbScanner, ie multi-page Bitonal, Group 4 TIFF images ideally at a minimum of 300 dpi.

Q: Can the system handle a single PDF/TIFF file containing multiple invoices?

A: No, whilst the solution can handle multi-page invoices, the image file needs to be one invoice transaction per file.

Q: How many network folders can be simultaneously monitored?

A: Up to 10 folders and 10 batch default types can be configured.

Glossary

A

accounts package: Version One packages can interface with third-party standard accounts and ERP packages.

annotation: A note added to a document in DbArchive. Like adding a sticky not to a paper document, the original document is not changed.

authorisation route: Documents can proceed through a number of approval states to be fully authorised. The authorisation route taken by a particular document follows predefined rules. Rules can use fields from the document, such as total value.

B

barcode: An optical machine-readable representation of data.

D

DbArchive: An enterprise-wide document imaging and management system, used to archive outgoing and incoming documents, and retrieve them quickly and easily.

DbAuthorise: Manages the authorisation of documents, such as supplier invoices requiring payment. Each document can pass through a series of authorisations before final approval.

DbCapture: Used to give a summary view of captures and lets you work with your documents. Captured information is shown and any problems with documents can be corrected.

DbConsole: Provides facilities for configuring and controlling the way DbScanner, DbQuery and DbWebQuery operate.

DbDeposit: Used to add extra document into the archive. Most documents are deposited automatically by printing invoices or scanning. DbDeposit allows you to add a file, such as a Word document or a PDF file, into the archive.

DbFax: Provides facilities for sending electronic documents by fax. Administrators can use DbFax to view fax requests on the server, edit or view system files, and take appropriate actions.

DbForm: Used to process source files (typically a print spool file), reformat them and pass the new layout to an output program. Output programs can convert the file to a different format, eg tiff, pdf, xml, csv, and pass it to other programs, such as DbFax, DbArchive and DbMail.

DbLogin: The common module used to access each of the document management suite applications.

DbMail: Module handling email integration, used to generate and send emails from your business applications. For example, send remittance advices directly to your suppliers via email, instead of printing, faxing or posting.

DbMailAdmin: The administration component of DbMail used to enter the information and select options to run DbMail how you want.

DbQuery: Provides facilities for finding and retrieving your archived and stored information.

DbScanner: Provides facilities for scanning paper documents and processing the images.

DbTrans: Before you can edit or delete a table you must unload the document data stored in it, and export it to another file.

DbWebquery: Provides facilities for retrieving the stored information through a standard web browser interface. Similar to DbQuery in use.

DbWebRetrieve: Uses a web browser to find and work with documents in your archive (DbArchive). You can search by a number of methods.

document profile: Document profiles control a set of scanning properties. It controls aspects to the scanning process such as whether barcodes are used, the OCR environment (if OCR is being used), how look-ups work for the document tags and so on. Each document profile is connected to a document type.

document type: A document type is created for each sort of document that is scanned. For example purchase orders, purchase invoices, goods received notes, etc.

duplex: Double-sided document printing.

E

EDI: Electronic Data Interchange — electronic exchange of transaction documents with your trading partners.

ERP: Enterprise Resource Planning — software applications that integrate internal and external management information across an organisation, embracing finance/accounting, manufacturing, sales and service, CRM, etc.

event log: A record of all the activity on your system — may be used by the support team to help diagnose any problems.

G

green meter: A package used to quantify the environmental savings from using the software in terms of CO2 emissions and tree equivalents.

GUI designer: Graphical tool available to produce the .def file. Input spool files can be mapped and output form definitions can be generated from scanned input files, data can then be placed on the output form.

I

import batch: Used if documents arrive through Electronic Data Interchange (EDI) or are scanned using another software package.

J

JPG: Graphic file format, uses 'lossy' compression, to give smaller file sizes.

K

Kofax: Kofax software is the image-processing engine used with DbScanner.

M

mail: Applications can use your email client to send a document from one user to another.

manual recognition: Sometimes a document cannot be OCRed properly, either because it was poor quality or hand-written. You are shown a warning for these fields and you can edit the values manually.

O

OCR: Optical Character Recognition — the electronic translation of images of text (usually captured by a scanner) into editable text. OCR is not an exact science and requires human validation to make sure the results are acceptable. Think in terms of around 90% accuracy, depending on fonts, size, quality of original, etc.

S

scanner: A device that optically scans images, printed text or handwriting and converts it to a digital image.

T

TIFF: Graphic file format, uses 'lossless' compression, giving larger file sizes.

V

Version One: Award winning document management company. Authors of the document management suite of applications, saving time, improving efficiencies and improving green credentials.

Index

A

about DbCapture23
about DbCapture client.....26
about DbCapture export24
about DbCapture interactive client31
about help19
accept document30
accepting documents24
adding a desktop shortcut 7
adding lines to grid45
address18
admin client.....26
anchors 23, 40
application suite 2
authorisation..... 2

B

backups17
best match12
business security17

C

capture properties 52
capture queue.....23
capture templates.....36
capturing data..... 2
capturing data lines45
change environment19
change language19
change password.....19
change user.....19
check folders.....62
choosing anchors.....40
choosing fields 9
clear search term 9
client administration26
common features 7
configuring PDF module.....59
contact us.....18
contents tab12
copyright.....18

correcting data.....31
creating a shortcut 7
creating anchors40
creating and using anchors.....40
customising columns..... 9
customising fields..... 9
customising the suite 7

D

data capture 2
data find 9
data from images58
data security.....17
data templates.....36
DbCapture admin context menu30
DbCapture admin right-click30
DbCapture client26
DbCapture export.....24
DbCapture interactive client.....31
DbCapture introduction23
DbCapture line level44
DbCapture properties.....52
DbLogin15
define document columns.....45
delete document30
deleting documents31
desktop shortcut 7
disaster recovery.....17
displaying anchors.....31
document accept.....30
document anchors.....40
document authorisation..... 2
document data capture 2
document fields.....36
document find..... 9
document identifier.....40
document identifiers37
document line templates45
document lists26
document management suite2, 7
document on hold.....23

document on-hold30
document properties..... 31, 52
document queue.....30
document tags23
document templates36
documents to be checked31
documents to check.....23
dot matrix bias.....23

E

editing capture data.....31
email contact18
environment19
example screens12
export captures24
extract line data 23, 31
extract lines.....45
extract tags 23, 31
extracting data manually45
extraction processing31

F

FAQs - PDF62
feedback18
field chooser 9
fields that move40
file backups17
filter by user 23, 31
filtering documents..... 9
find date..... 9
finding document fields36
finding documents 9
finding help.....12
finding us18
fixing captures31
floating anchors.....40

G

getting in touch.....18
glossary12
go to menu.....31

H

handwritten documents..... 31
help about..... 19
help contents..... 12
help index 12
help version..... 18
hold document23, 24, 30

I

identifying document fields..... 36
import PDF scans 58
import TIFF scans..... 58
importing image files 58
importing PDF..... 57
importing PDFs 62
importing TIFF 57
importing TIFFs..... 62
index tab..... 12
info address..... 18
installation 2
installing PDF module..... 59
interactive client..... 31
invoice filter..... 23
issue with help..... 18

L

language..... 19
layout options..... 9
line level capture..... 44, 45
line level template..... 44
live environment 19
login 15
looking for a help topic..... 12

M

manual data extraction 45
manual line changes 45
menu options..... 7
modules..... 2
monitoring folders 57
monitoring network folders..... 62
moving columns..... 9

moving fields40
moving user's documents26
multi-page invoice imports 62

N

named user's documents23

O

OCR anchors40
OCR line level44
on hold reason30

P

page width view31
password19
PDF and TIFF import57
PDF configuration59
PDF import 58, 62
PDF import module57
PDF module57
PDF module installation59
PDF tag information59
personalising 9
phone number18
polling directories57
pop-up screen12
primary identifier40
problems with help18
processing lists26
product family 7
properties - captures52
purchasing cycle23

Q

queries18
queue captures23
queues30

R

recently used menu list 7
reinstate document30
release number18
remove document30

remove search term 9
reprocess document30
revalidate document30
rubber stamps23

S

sales cycle23
saving user data 23, 31
scan anchors 40
scan quality23
scan templates36
scanning document lines44
scanning PDF57
screen display 9
screen shots12
search date 9
search fields 9
search operators 9
search tab12
searching help12
secondary identification 23, 31
secondary identifier40
securing data17
select fields 9
select language19
server connection31
show on hold reason30
Show/hide screen12
signing in15
software version18
starting the software 7
sub-queues26
system administration26
system security17

T

tags 23, 36
technical terms12
templates 36, 45
test environment19
TIFF import57, 58, 62

to be checked23
tools 2

U

understanding document identifiers37
update password19
user change.....19
user document lists.....26
user login15
user password.....15
username15
using anchors.....40
using help.....12
using line level capture.....45
using the software 7

using the windows 9

V

version number19
version one18
view document.....30

W

which tool to use..... 2
wilmslow18
window customisation 9

Z

zooming in31
zooming out31