



DataGateway User Manual

Innovative Technology

DataGateway User Manual

Version 8a986b7 (master), 09 May 2013

Copyright © 2013 William Matthew Limited. All rights reserved.

Acknowledgements.

Microsoft and Windows are registered to Microsoft Corporation.

HP and PCL are trademarks of Hewlett Packard.

PowerEdge is a trademark of Dell Corporation.

VMware is a trademark of VMware, Inc.

Linux is a registered trademark of Linus Torvalds.

Prescribe is a registered trademark of Kyocera-Mita Corporation.

E&OE

Table of Contents

About this manual	vii
1. Intended Audience	vii
2. Software Licence and Copyright	vii
1. Introduction	9
2. Getting Started	11
2.1. Platform set-up	11
Setting up a Server-based DataGateway	11
Setting up a Virtualized DataGateway	11
2.2. Accessing the console	48
Accessing the console on Tempest appliances	48
Accessing the console on Virtualized appliances	48
Accessing the console on Server-based appliances	49
2.3. IP address configuration	49
Console IP address configuration	49
2.4. Accessing the DataGateway user interface	50
DataGateway users	50
3. System Maintenance	51
3.1. Local storage	51
4. Disaster recovery and backup procedures	53
4.1. Backup	53
Automated backup with WML's disaster recovery system	53
Static virtual machine backup procedure	53
Database backups	55
Making backups of DataGateway system settings	55
4.2. Recovery	55
5. DataGateway Configuration	57
5.1. System configuration	57
Network settings	57
System date and time	59
System timezone and locale	60
System logging	60
Email settings	64
Licence key	64
System passwords	65
System settings	66
Firmware	68
Status and Control	69
6. Queue Management	71
6.1. Queue management interface	71
Print user actions	72
Print admin controls	73
Classes	75
6.2. Queue configuration	76

Data input	76
Overlay templates and logic profiles	83
Archiving	84
Failover	87
Alerts (Queue Monitoring)	88
PPD driver	90
Driver options	91
Output	92
Queue info	97
6.3. Queue configuration: Example workflow	97
6.4. Jobs	102
6.5. Platform settings	103
Cups Control	103
Failover	103
Edit cupsd.conf	103
Edit mime.types	103
Edit mime.convs	103
7. Template Management	105
7.1. Introduction	105
7.2. Overlay elements	107
7.3. Overlay templates	108
7.4. Overlay template list	108
7.5. Overlay template manager	109
Saving changes to the template	110
Template management	111
Template settings	112
The canvas	116
The overlay elements user interface	118
Testing print data	129
7.6. Overlay Templates: Example workflow	130
Create a print queue	130
Create PDF forms	130
Create and name a new template	131
Upload and activate PDF forms	132
Add text overlays	134
Review template settings	136
Upload and test	136
Bind the template to the queue	137
7.7. Using structured input data	138
XML input data	139
CSV data	144
Jetforms format input data	145
Template special keys	146
7.8. Using data groups	146
Letter with variable content	146

Multiple page invoice	151
8. Logic Profiles	161
9. Document Content Control	167
9.1. Overview	167
9.2. Configuration	168
Profile configuration	168
Page types	170
Document types	173
DCC actions	176
9.3. Testing the configuration	192
Example document analysis	192
9.4. Using document content control	194
Setting up a queue for document content control	194
Held documents	195
Output documents	196
10. File Manager	199
10.1. Form Files	200
10.2. Log Files	200
10.3. Archive Files	200
10.4. PPD Files	200
10.5. All User Files	200
Print Diagnostics	200
11. Print Reports	201
11.1. Viewing reports in the web interface	203
11.2. Reports with a pair of DataGateway devices	204
12. Appliance Manager	207
12.1. The Appliance Manager push	207
12.2. The user interface	207
12.3. Appliances and groups	208
Populating the appliance list	209
Appliance status	210
Appliance report	211
Managing pushes	213
12.4. Stored configurations	215
12.5. Stored firmware	216
12.6. Appliance Manager for Disaster Recovery	216
12.7. Appliance Manager Networking	217
13. Electronic Document Delivery	219
13.1. Overview	219
13.2. Configuration	220
Core and per-queue settings	220
Message templates	222
13.3. Using electronic document delivery	224
Setting up a queue for electronic document delivery	224
Print data for electronic document delivery	225

Delivery of electronic documents	225
Using the customer support interface	227
13.4. Web portal for electronic delivery	229
Setting up the electronic document delivery web portal	229
14. WML Secure Forms	235
15. Kyocera Prescribe Emulation	239
15.1. Basic Defaults	240
15.2. Paper Cassettes	240
Using Digital Paper Stock	241
Using PostScript Paper Selection Tags	241
15.3. Persistent State	241
16. Handling output from CDL Strata	243
16.1. Extracting CSV from Strata PostScript output	243
16.2. Strata input queue	243
16.3. Strata output queue	244
16.4. CSV data	244
16.5. Errors	245
17. Output Batching	247
17.1. Output batching introduction	247
17.2. Queue configuration for output batching	248
17.3. Output batch list	250
Releasing output batches	251
Downloading output batches	252
17.4. Output batch templates	256
Header and trailer templates	256
Document template	259
17.5. Re-printing spoiled documents	261
Setting up archiving on the final queue	261
Re-printing the document	262
18. Print Outsource: Sender Site	265
18.1. Print Outsource Introduction	265
18.2. Deployment	265
18.3. External Network	265
18.4. Internal Network	268

About this manual

This manual describes the use of WML DataGateway software.

1. Intended Audience

This manual is aimed at IT administration staff who will be involved with maintaining and configuring the DataGateway and at print production staff who will be reviewing reports.

2. Software Licence and Copyright

The Software is copyrighted and all rights reserved by William Matthew Limited, 22 Haxey Lane, Haxey, Doncaster, South Yorkshire, DN9 2NE. UK. Email Info@wmltd.co.uk

Important

Before using the WML DataGateway Software please read the licence agreement carefully. The right to use this software is granted only if the Customer agrees to the terms of this Licence. If you do not agree to the terms of this Licence, please return the product unused.

William Matthew Limited considers that installation of the Software is your acceptance of our terms and conditions of the license.

1. The Software has been developed for use with the WML DataGateway; The Licence is to be used without any time limit and may be subject to change.
2. The Software may not be duplicated, copied, resold, or re-branded without the written consent of William Matthew Limited.
3. It is the responsibility of any user, reseller, distributor, OEM or other third party to protect the rights of William Matthew Limited against violation of this Software Licence.
4. Some portions of The Software have been developed by third parties and are used by WML under the terms of the GNU General Public Licence, as well as other open source and free software licences. The source code to these portions of The Software are available to WML's customers under the same terms under which WML uses them. WML will uphold and protect the copyright of the original author of these portions of The Software.
5. William Matthew will consider any attempt to duplicate the functionality of any application provided by the software to constitute a breach of software licence, and may instigate legal proceedings to protect the copyright of the DataGateway software and William Matthew Limited.
6. This Licence and the Software product may be transferred to a third party provided the third party agrees to all the terms of this Licence and that the transferring party retains no copies of the Software.
7. William Matthew Limited reserves the right to terminate this Licence upon breach.

About this manual

8. In the event of a termination, all copies of the software must be destroyed or returned to William Matthew Ltd
9. Any disputes in respect of the Software Licence will be considered a breach of copyright and subject to English Law.

Chapter 1. Introduction

WML DataGateway is a *transactional and variable data printing solution*. It is available as an embedded hardware appliance or as a virtual appliance for VMware or Linux KVM.

DataGateway is an ideal solution to *securely outsource print fulfilment* from multiple locations to a central production facility. Outsourced print processing can be achieved in *real-time*. No time is lost building up and distributing batch files, which significantly extends the working day.

This makes DataGateway ideal for the production of documents which must comply with minimum time-to-customer regulations.

Outsourced print is never held in non-volatile storage at the print fulfilment site, which makes it *suitable for use with documents which fall under the Data Protection Act*.

DataGateway can incorporate *document content control* functionality. Documents are parsed by the DataGateway's logic engine and are then evaluated for completeness and correctness according to a ruleset developed by WML in conjunction with the customer. Incorrect or incomplete documents are held in a database pending action from print managers.

DataGateway provides a *real-time document audit trail* and daily and monthly print volume reports.

DataGateway can accept data from your mainframe, Unix, Windows and Linux application servers using Berkeley LPD protocol, IPP protocol and simple TCP raw-data transfer. It is also able to collect print jobs from Windows (SMB/CIFS) and Unix (NFS) network-shared "hot folders". In addition, FTP can be used to send input data, enabling the DataGateway to be used in conjunction with multi-function printers to process scanned images.

DataGateway can masquerade as multiple printers by assuming an IP address for each printer. Using this feature, it is possible to use DataGateway to replace a number of desktop printers with a single production printer requiring no change to the application which generates the print jobs.

Print jobs are output, via a print driver, to your destination printers. LPD, IPP or raw protocols can be used. Archiving or emailing of jobs (in PDF format) is available.

DataGateway has the versatility to help you to implement new print functionality and streamline your print factory. It reduces errors, increases print security and helps you to control your print costs.

Chapter 2. Getting Started

In this chapter, you will be guided through the initial set-up for the WML DataGateway. DataGateway is available in three platform configurations:

- *Tempest DataGateway* is an appliance form-factor device, based on an AMD Geode LX processor.
- *Server-based DataGateway* is a Dell or Intel server appliance, for situations where faster processing or local (hard disk) storage is required. Rack servers provide more network interfaces than the Tempest platform, which can be a requirement in enterprise networks.
- *Virtualized DataGateway* is a VMware virtual machine image of the software. Virtualized DataGateway can be delivered as a pure image to install on the customer's own VMware infrastructure, or on server hardware delivered and maintained by WML. Virtual print appliances are also available in a format suitable for running under Linux KVM (as used in Ubuntu Linux).

2.1. Platform set-up

In this section, you will find out how to set up the various hardware and virtualized platform configurations. The following sections describe set-up tasks that are common to all configurations.

Setting up a Server-based DataGateway

Important

The exact choice of server equipment will vary depending on the expected workload for DataGateway. For this reason, it's not possible to give a comprehensive setup guide for a Server-based DataGateway here.

The Server-based DataGateway will consist of a one or two unit rack-mount server, with a suitable number of processors, RAM and disk storage for the intended application and print throughput. It will be delivered as a single unit, complete with a rack-mount kit.

The Server-based DataGateway will be delivered fully configured and ready to run. Power the device on and direct your web browser to the IP address provided by WML staff.

Setting up a Virtualized DataGateway

Virtualized appliance disk files are distributed either as VMWare images (.vmdk files) or a Linux Kernel Virtual Machine (KVM) compatible images (.qcow2 files).

Follow the procedure given by the vendor of your virtualization infrastructure to create a new virtual machine using the Virtualized appliance disk file as the virtual disk.

Note

The Virtualized appliance disk file will be read from and written to. For this reason, it is advisable to keep a copy of the original disk file from which a fresh virtual machine can be created. Select a suitable location on your disk system to store your in-use disk files. This should be somewhere where it is unlikely that the files will be moved or interfered with externally.

Make a separate copy of the Virtualized appliance disk file for each new virtual appliance which you create.

Adding a second virtual disk for increased storage

The Virtualized appliance disk file provides up to 2.5 GB of space to store settings, DCC profiles and PDF forms. This is a sensible amount of storage for systems which do not need to archive locally, store log files locally for long periods of operation or run an internal database.

In practice, DataGateway systems typically require more storage as they will, in most cases, run an internal database (MySQL), archive document data and store log files internally. PdfPlus and PrePrint systems usually work well with the default disk file.

All WML Print Platform appliances will recognise the presence of a second virtual hard disk and make use of this disk to store all persistent data in place of the persistent partition of the Virtualized appliance disk file.

When the second disk is first detected during a boot of the Virtualized appliance, the system tests it to see if it already contains the persistent files from a WML Print Platform system. If it does contain files, it is attached as the persistent storage, and boot up continues as normal. If it is empty, then the boot up procedure first copies all the files from the Virtualized appliance disk file's persistent partition to the second disk, then completes boot up using the second disk as persistent storage.

To make use of a second disk, simply create a new disk in VMware or KVM, create a Linux ext3 filesystem on the entire disk (that is, don't create any partitions on the new disk), ensure it is attached to your virtual machine as the second disk and reboot your appliance. Ideally, the new disk should appear as a SCSI disk to the operating system. It should appear internally on the Linux-based WML Print Platform system as the device `/dev/sdb` for correct results.

VMware infrastructure options

VMware provides a number of different infrastructure options, including VMware workstation, VMware ESX and VMware vSphere/ESXi.

A Virtualized DataGateway should work on any type of VMware infrastructure, from the free VMware Server, through to VMware vSphere systems, however, you do need to choose to download the correct vmdk disk image from our servers.

VMware virtual disks, pre-installed with DataGateway software are available in two formats. The first is a growable disk provided as a single file with a name such as:

Getting Started

```
wmlpp-dg-vmware-sata-tree3-20130228_1502.vmdk
```

This format can be used as the virtual disk for virtual machines created on VMware Workstation and VMware Server, but not recent versions of VMware ESXi/vSphere.

If you are using VMware ESXi or vSphere, you will need to download the ESX-compatible VMDK. This is provided as a zip file with a file name suffix **.esx.zip**. The zip contains two files, a `.esx.vmdk` disk descriptor file and a 2.5 GB pre-allocated `.esx-flat.vmdk` file which contains the disk data. For example:

```
wmlpp-dg-vmware-sata-tree3-20130228_1502.esx.vmdk  
wmlpp-dg-vmware-sata-tree3-20130228_1502.esx-flat.vmdk
```

Upload both of these `.vmdk` files to the data store of your VMware infrastructure before creating a new virtual machine.

Note

It is possible to use the ESX-compatible VMDK files on VMware Workstation.

VMware Server walk-through

This walk-through demonstrates the set up of a Virtualized appliance on VMware Server version 2.0.2.

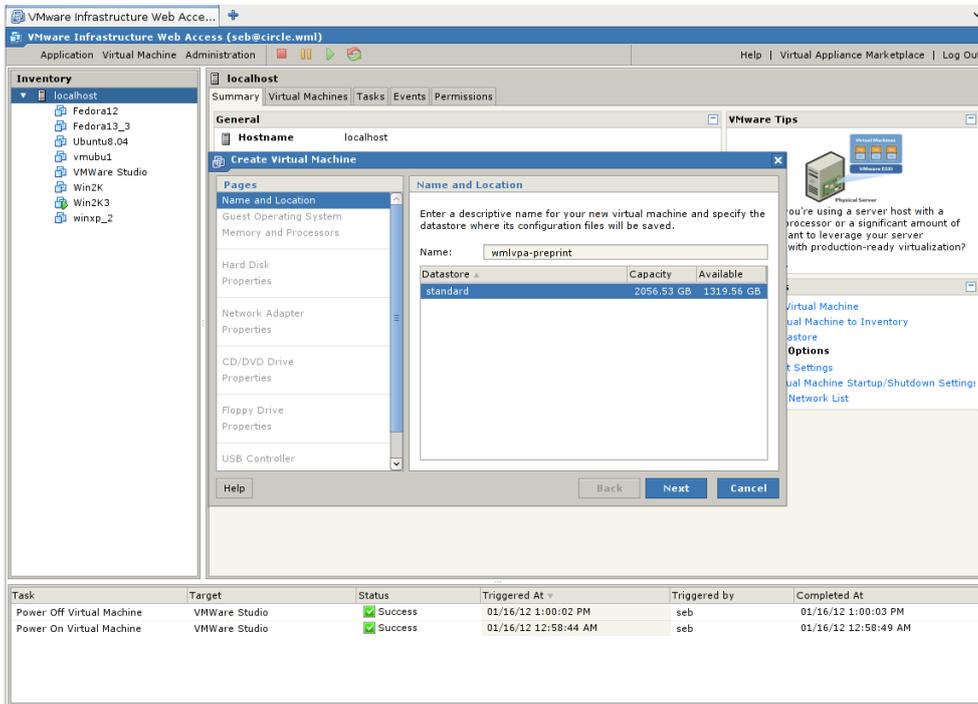
Note

It is assumed that you have VMware Server version 2.0.2 already installed on your PC. For help installing VMware Server, please refer to the VMware website.

VMware Server 2.0.2 can be obtained both for Linux and for Windows platforms; our appliances have been verified to work correctly on both.

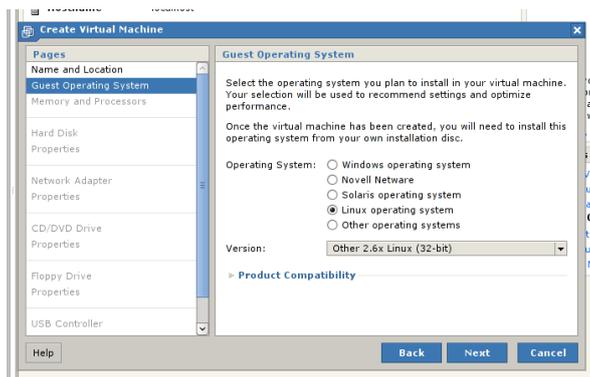
Creating a new Virtualized appliance is carried out in the VMware Infrastructure Web Access interface. This is accessed via a web browser on the system on which VMware Server was installed. Typically, it can be accessed at the address **<http://yourcomputer:8333/>**

Figure 2.1. Creating a Virtual Machine in the VMware Infrastructure Web Access Interface.



In VMware Infrastructure Web Access, click **Virtual Machine > Create Virtual Machine**. The **Create Virtual Machine** dialog will open (Figure 2.1). Assign a name for your new appliance. Here it has been called “wmlvpa-preprint”. Click **Next**.

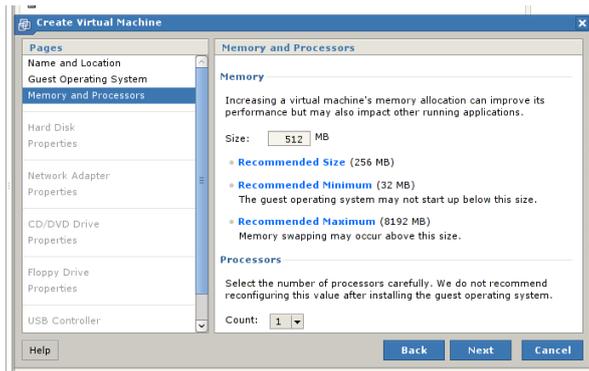
Figure 2.2. Specifying the operating system for the guest appliance.



The next window (Figure 2.2) allows you to choose which operating system your target appliance will be running. All WML appliances run a 32 bit Linux operating system and so

you should choose **Linux operating system** for **Operating System** and **Other 2.6x Linux (32-bit)** for **Version**. Click **Next**.

Figure 2.3. Setting the amount of RAM and number of processors for the virtual appliance.



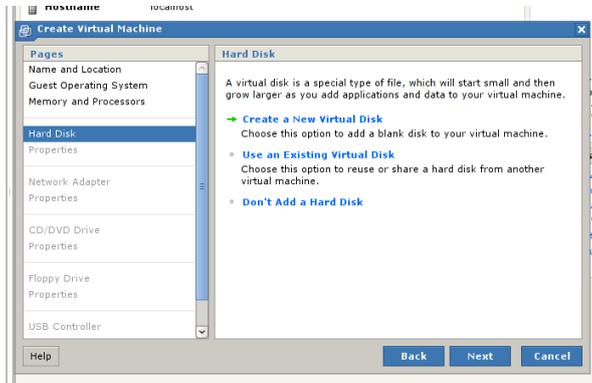
This screen (Figure 2.3) allows you to specify the initial amount of RAM and processor resource to assign to your virtual print appliance. Specify a minimum of 256 MB of RAM. Some applications may require more memory to run correctly, but 256 MB is sufficient for simpler tasks and short documents.

Assign the desired number of processors for the virtual appliance. A single processor may be sufficient for simpler applications. Assigning two processors to the system will improve the responsiveness of the web interface when the print processing load is high.

Note

The amount of RAM may be altered after the virtual appliance has been created. Altering the number of processors is not recommended by VMware, but it can be done. WML appliances will safely handle a change in the number of available processors.

Figure 2.4. Adding a hard disk to the virtual appliance.



In the next window (Figure 2.4) you have the option to **Create a New Virtual Disk**. This would create an *empty* disk, which would then need to have the operating system installed (perhaps from a dvd or cdrom). Because you are using a pre-populated disk file supplied by WML, you should click on **Use an Existing Virtual Disk**, which will open the next window, shown in Figure 2.5.

Figure 2.5. Setting the hard disk properties.

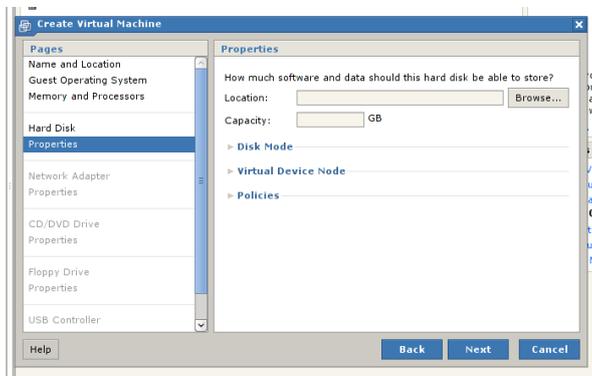
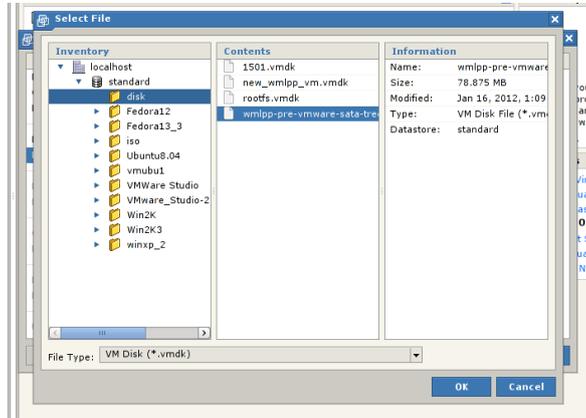


Figure 2.6. Selecting the WML Virtualized disk file.



On the **Properties** window (Figure 2.5), click **Browse** and browse to the location where you saved your .vmdk disk image (See Figure 2.6) then click **OK** (it's not necessary to change **Disk Mode**, **Virtual Device Node** or **Policies** from their default values).

Warning

VMware Server will *not* make a copy of this disk image file; it's up to you to keep a copy of the original disk image. The disk image file will be read from and written to, and may increase in size to a maximum of approximately 2.5 GB.

Figure 2.7. The Network Adapter window.

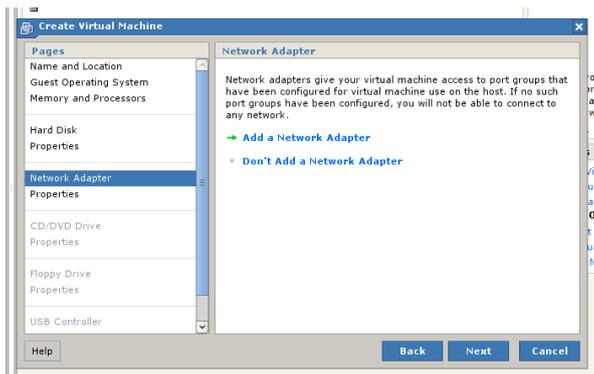
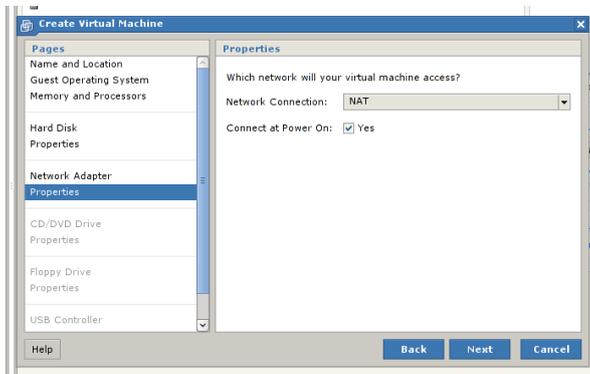


Figure 2.8. The Network Adapter properties window.



In the **Network Adapter** window (Figure 2.7), click **Add a Network Adapter**. In the properties window for the network adapter (Figure 2.8), you can select different kinds of network. For a most systems, you are likely to need **Bridged** networking, which allows your virtual appliance to join your physical network as if it were a real machine. **NAT** (Network Address Translation) networking is unlikely to be suitable, as the WML Virtualized appliance acts as a server. For test installations, **HostOnly** may be a suitable choice. In this case, the virtual machine exists on an internal, virtual network.

A full discussion of networking your virtual machines is outside the scope of this document. If you are unsure of which type of network connection to choose, contact your network administrator or your networking consultant or supplier.

Leave **Connect at Power On** checked, and press **Next**.

Figure 2.9. The CD/DVD Drive window.

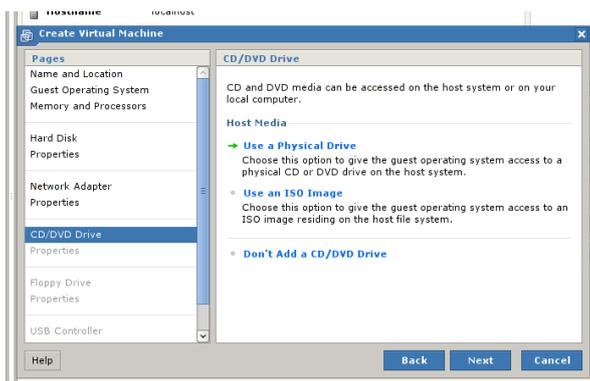


Figure 2.10. The Floppy Drive window.

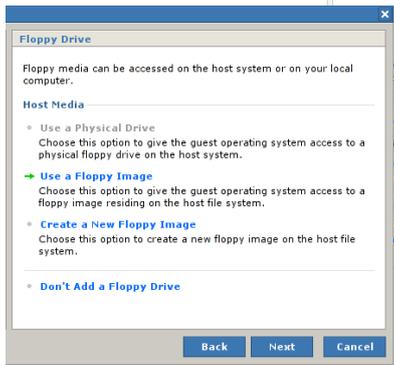
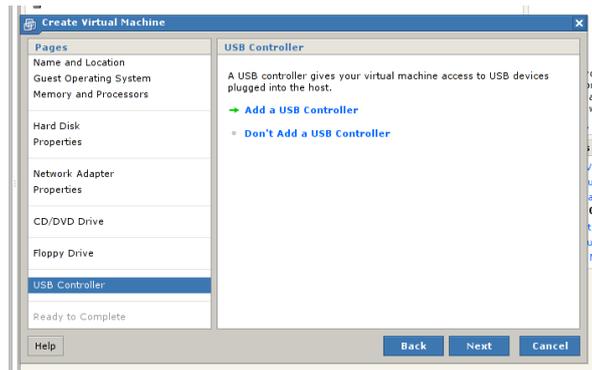
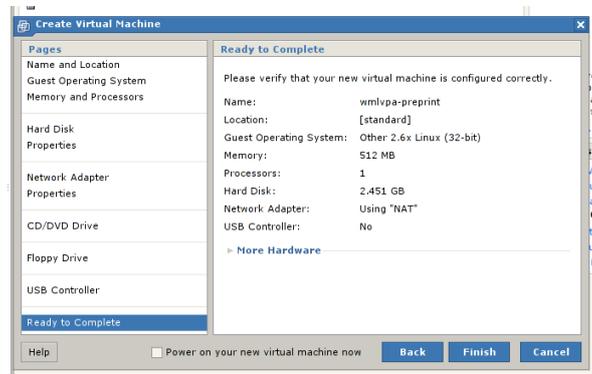


Figure 2.11. The USB Controller window.



On the CD/DVD Drive window (Figure 2.9), click **Don't Add a CD/DVD Drive**. Likewise, on the floppy drive (Figure 2.10) and USB Controller (Figure 2.11) windows, click "Don't Add..."

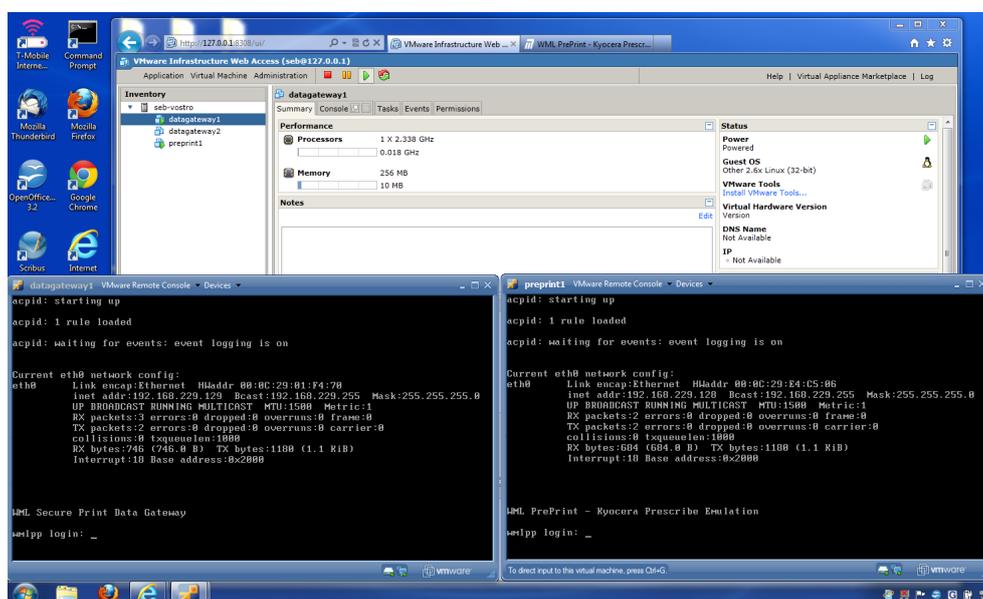
Figure 2.12. Ready to Complete - a summary of your new virtual appliance.



Click **Next** on the USB Controller window and you will arrive on the summary window (Figure 2.12). When you click **Finish**, your Virtualized virtual appliance is ready to use.

You can access the console of your virtual machines via the VMware Infrastructure Web Access interface. Figure 2.13 shows a Windows desktop with two running Virtualized appliances; a DataGateway and a PrePrint.

Figure 2.13. Two VMware virtual consoles; one is a WML DataGateway, the other is a WML PrePrint appliance.



We hope this walk through for VMware Server is useful. Other VMware infrastructure options will differ from VMware Server, but the process will be similar to the one shown above.

VMware Workstation walk-through

This walk-through demonstrates the set up of a Virtualized appliance on VMware Workstation version 9.0 running on Microsoft Windows 8.

Note

It is assumed that you have VMware Workstation version 9.0 already installed on your PC. For help installing VMware Workstation, please refer to the VMware website.

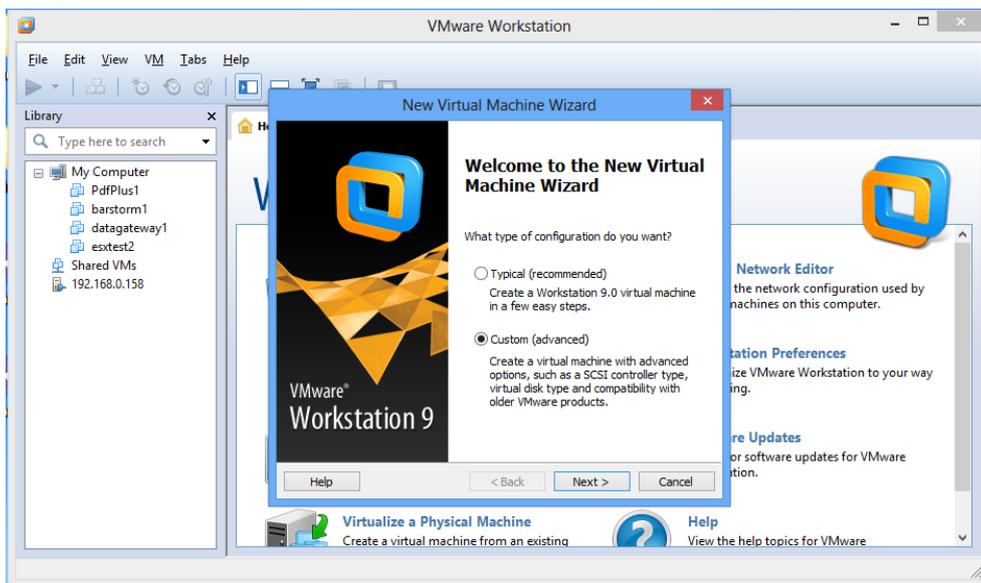
Creating a new Virtualized appliance is carried out in the VMware Workstation user interface. This is accessed via an icon in your start menu or on the desktop.

Figure 2.14. The Home tab of the VMware Workstation user interface.



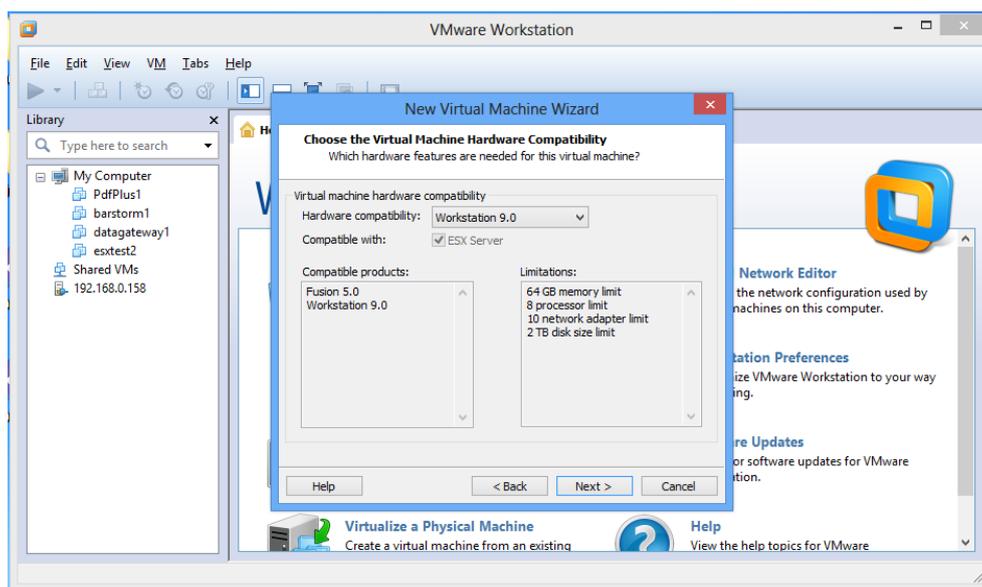
In the VMware Workstation Home tab, click **Create a New Virtual Machine**. The **New Virtual Machine Wizard** will open (Figure 2.15). Select **Custom (advanced)** so that it will be possible to specify that we want to use an existing, preformatted virtual disk. Click **Next**.

Figure 2.15. The “New Virtual Machine Wizard”.



The first window allows you to specify the hardware compatibility for the virtual machine. This may be a consideration if you wish to move your virtual machine to older versions of VMware Workstation or to other VMware infrastructure products. In our example, we leave it at its default of **Workstation 9**.

Figure 2.16. Choosing the virtual machine hardware compatibility.



As we will use a pre-existing virtual disk, containing a ready-to-boot operating system, select **I will install the operating system later** on the **Guest operating system installation** window (Figure 2.17).

Figure 2.17. Selecting the source for the operating system for the virtual machine.

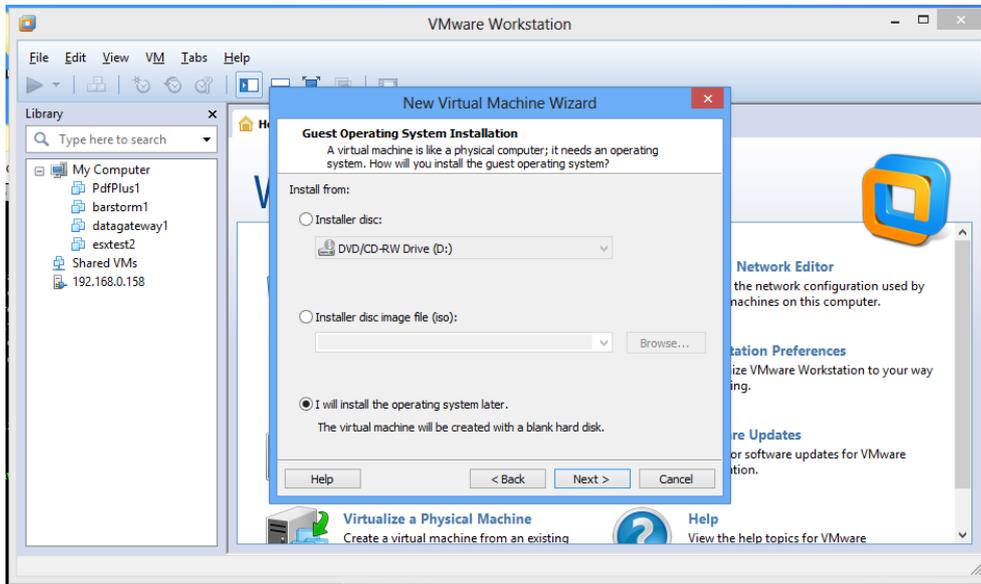
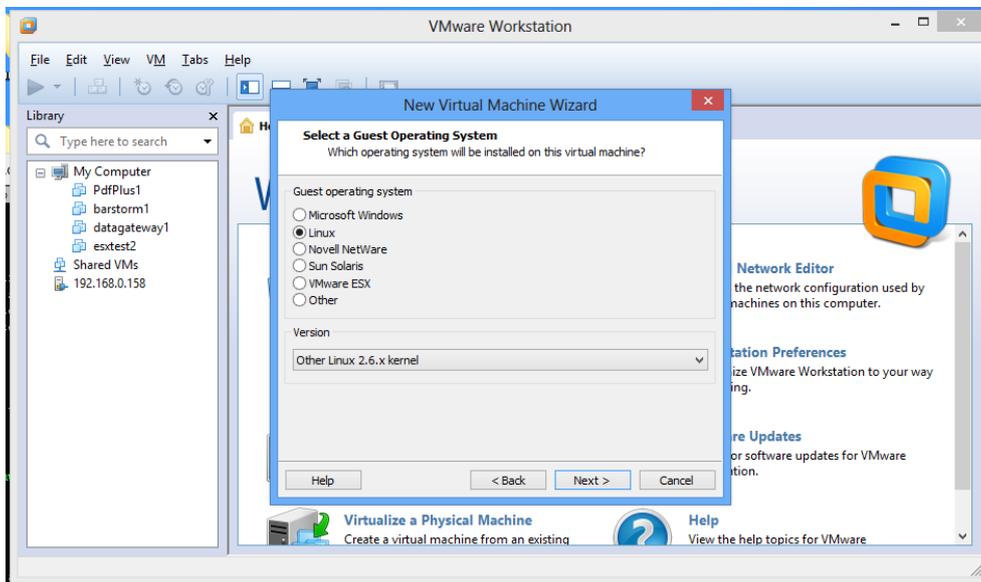


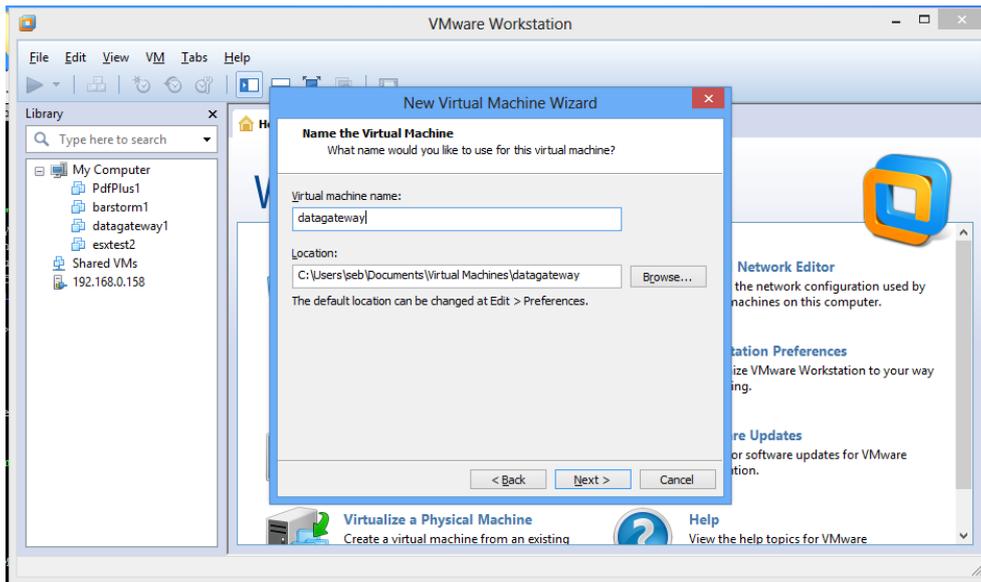
Figure 2.18. Specifying the operating system for the guest appliance.



The next window (Figure 2.18) allows you to choose which operating system your target appliance will be running. All WML appliances run a 32 bit Linux operating system and so

you should choose **Linux** for **Guest operating system** and **Other Linux 2.6.x kernel** for **Version**. Click **Next**.

Figure 2.19. Choosing the name for the virtual machine.



You can now choose the name for your virtual machine. Here, it has been named “data-gateway” (Figure 2.19). The virtual machine definition files will be created in the default location **Documents\Virtual Machines**.

Figure 2.20. Setting the number of processors for the virtual appliance.

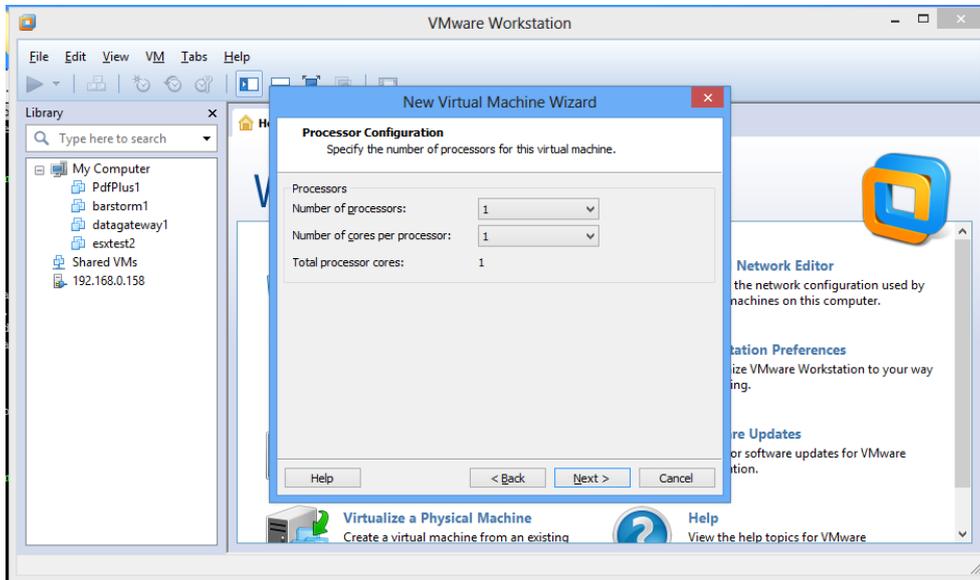
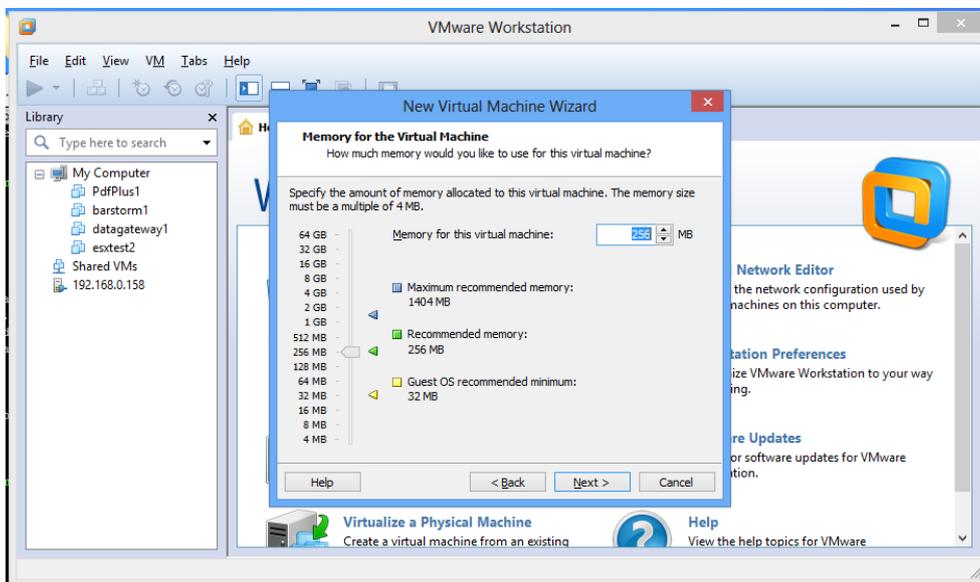


Figure 2.21. Setting the amount of RAM for the virtual appliance.



The next two screens (Figure 2.20 and Figure 2.21) allow you to specify the initial amount of RAM and processor resource to assign to your virtual print appliance.

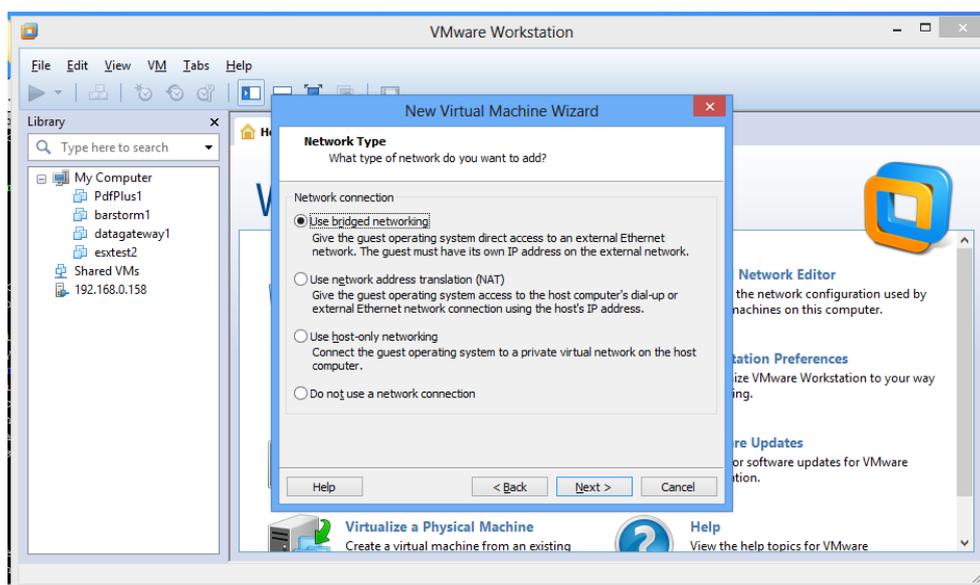
Assign the desired number of processors for the virtual appliance. A single processor may be sufficient for simpler applications. Assigning two processors to the system will improve the responsiveness of the web interface when the print processing load is high.

Specify a minimum of 256 MB of RAM. Some applications may require more memory to run correctly, but 256 MB is sufficient for simpler tasks and short documents.

Note

The amount of RAM may be altered after the virtual appliance has been created. Altering the number of processors is not recommended by VMware, but it can be done. WML appliances will safely handle a change in the number of available processors.

Figure 2.22. The Network Type window.



In the **Network Type** window (Figure 2.22) you can select different kinds of network. For a most systems, you are likely to need **Use bridged networking**, which allows your virtual appliance to join your physical network as if it were a real machine. **Use network address translation (NAT)** is unlikely to be suitable choice, as the WML Virtualized appliance acts as a server. For test installations, **Use host-only networking** may be a suitable choice. In this case, the virtual machine exists on an internal, virtual network.

A full discussion of networking your virtual machines is outside the scope of this document. If you are unsure of which type of network connection to choose, contact your network administrator or your networking consultant or supplier.

Figure 2.23. Selecting the virtual I/O controller type.

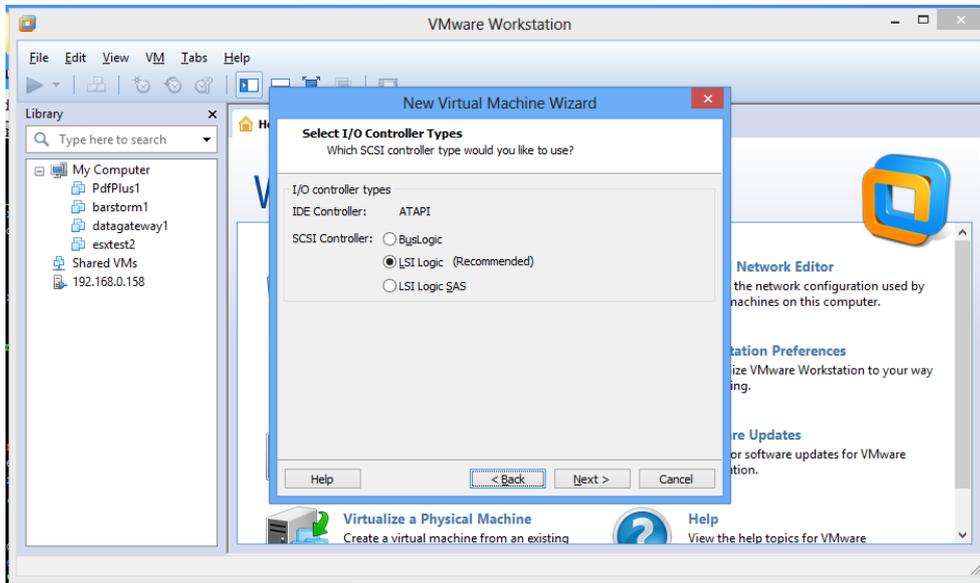
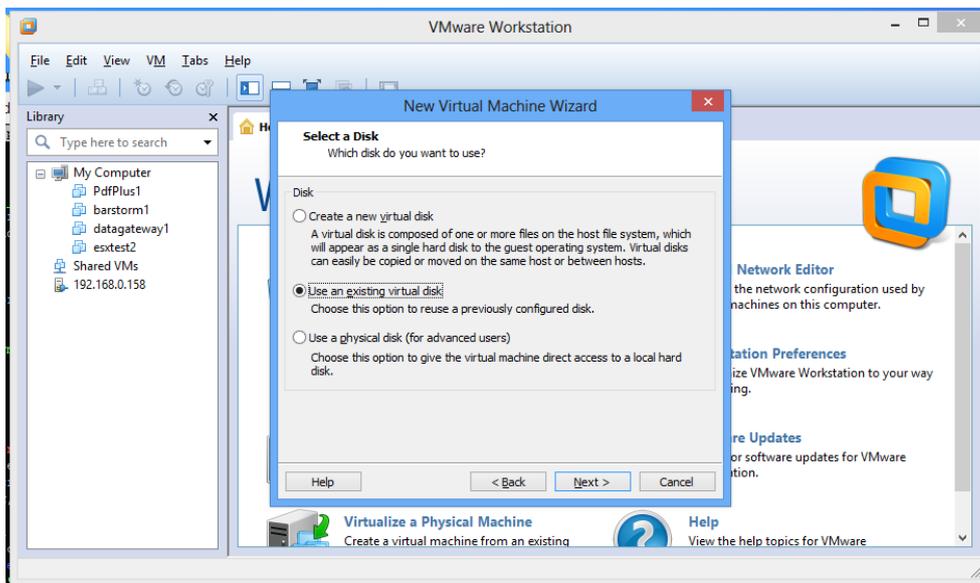


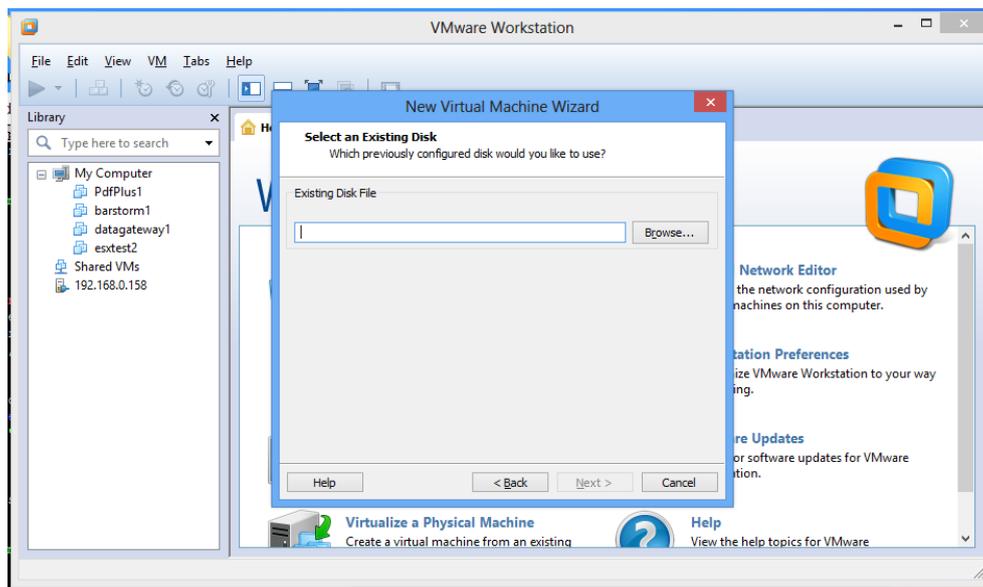
Figure 2.23 shows a screen allowing a choice of virtual disk I/O controllers. Leave the default **LSI Logic (Recommended)** checked.

Figure 2.24. Selecting a hard disk for the virtual appliance.



In the next window (Figure 2.24) you have the option to **Create a new virtual disk**. This would create an *empty* disk, which would then need to have the operating system installed (perhaps from a dvd or cdrom). Because you are using a pre-populated disk file supplied by WML, you should click on **Use an existing virtual disk**, which will open the next window, shown in Figure 2.25.

Figure 2.25. Selecting an existing disk.



Browse to the vmdk disk image (Figure 2.26). You will probably see a pop-up asking if you wish to convert the existing virtual disk to a newer format (Figure 2.27). You can choose to do this or not, as you prefer; the virtual machine will boot correctly in either case.

Warning

VMware Workstation will *not* make a copy of this disk image file; it's up to you to keep a copy of the original disk image. The disk image file will be read from and written to, and may increase in size to a maximum of approximately 2.5 GB.

Figure 2.26. Selecting the WML Virtualized disk file.

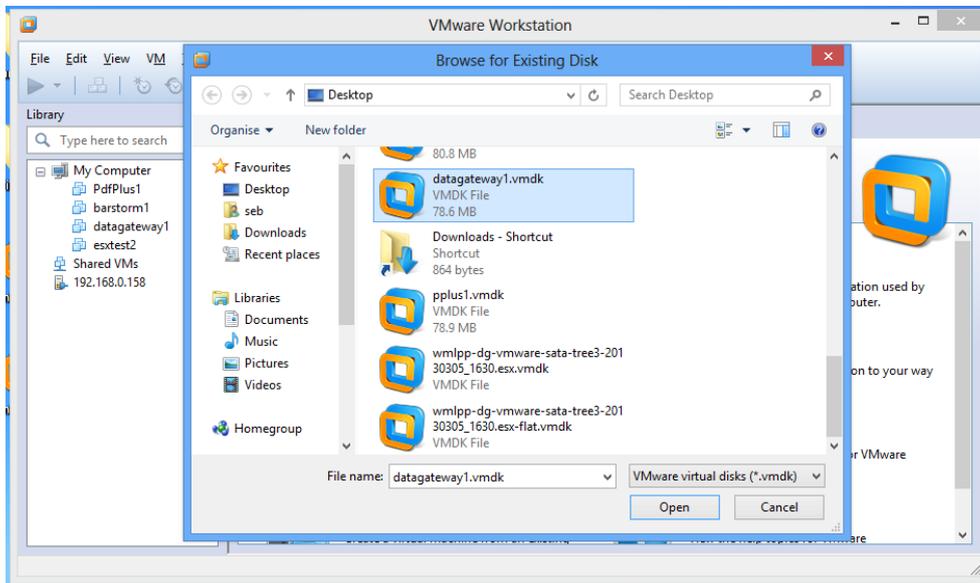
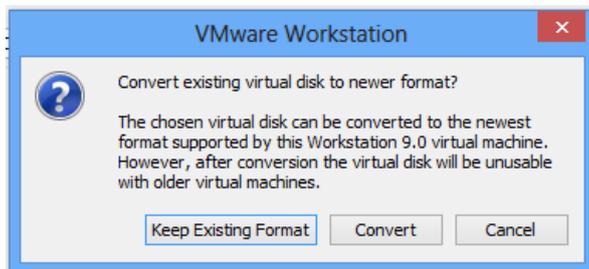
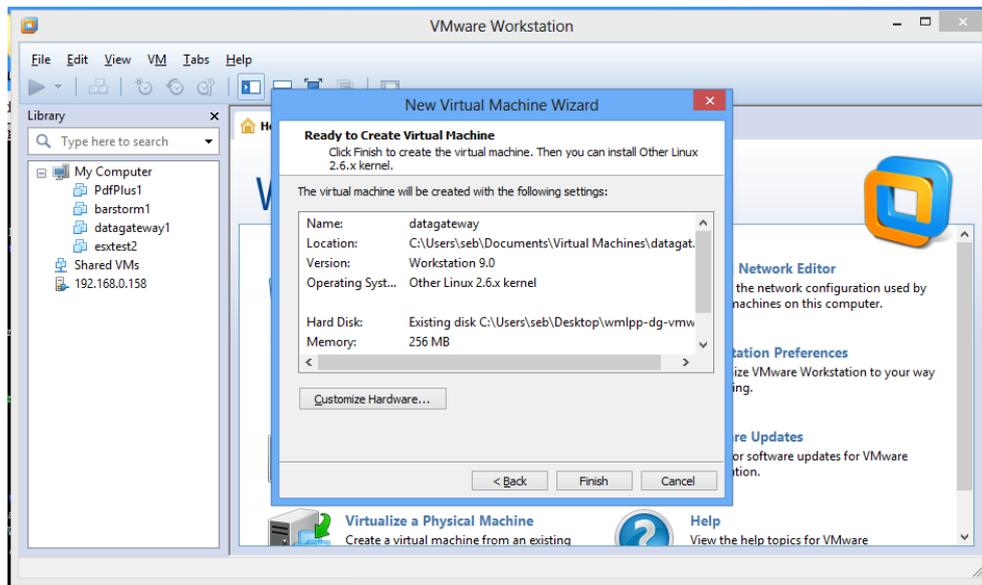


Figure 2.27. Converting to newer format.



Choose to keep the existing disk format, or convert it and press **Next**. You'll see the final window in the wizard (Figure 2.28). When you click **Finish**, your Virtualized virtual appliance is ready to use.

Figure 2.28. Ready to Create Virtual Machine - a summary of your new virtual appliance.

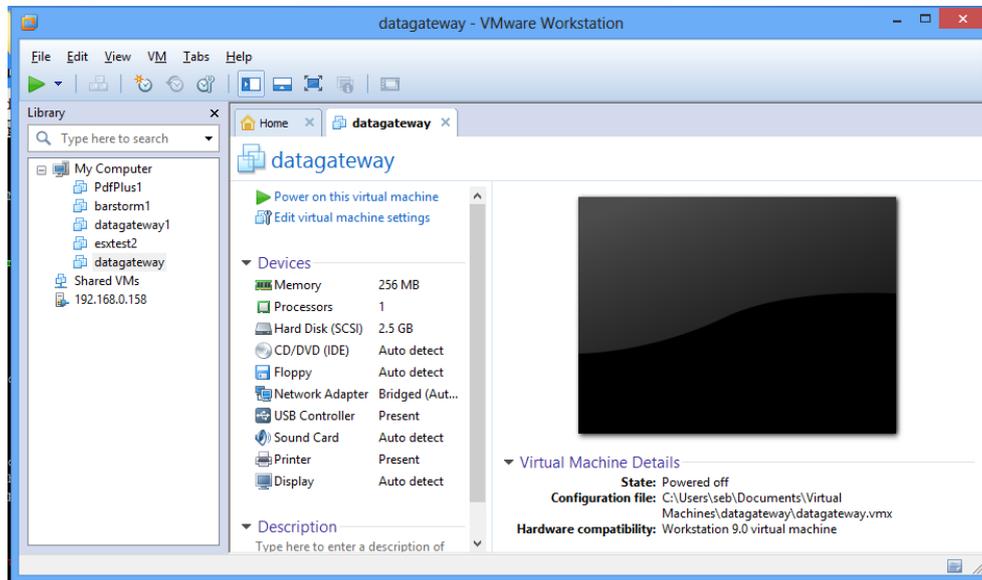


After creating your new virtual machine, a new tab will appear in the VMware Workstation UI. You can access the console of your virtual machine here (Figure 2.29). When you first boot your virtual machine, you will see a message about installing a guest operating system. You can simply click the button labelled **I Finished Installing**.

Note

There is no need to install VMware tools on a WML Virtualized. VMware tools enhances the user experience when using a mouse within a graphical environment. The DataGateway has no graphical console and hence VMware tools is not used.

Figure 2.29. The new appliance tab.



We hope you found this walk-through for VMware Workstation useful. If you compare this with the VMware Server walk-through, you will see the process has many common features.

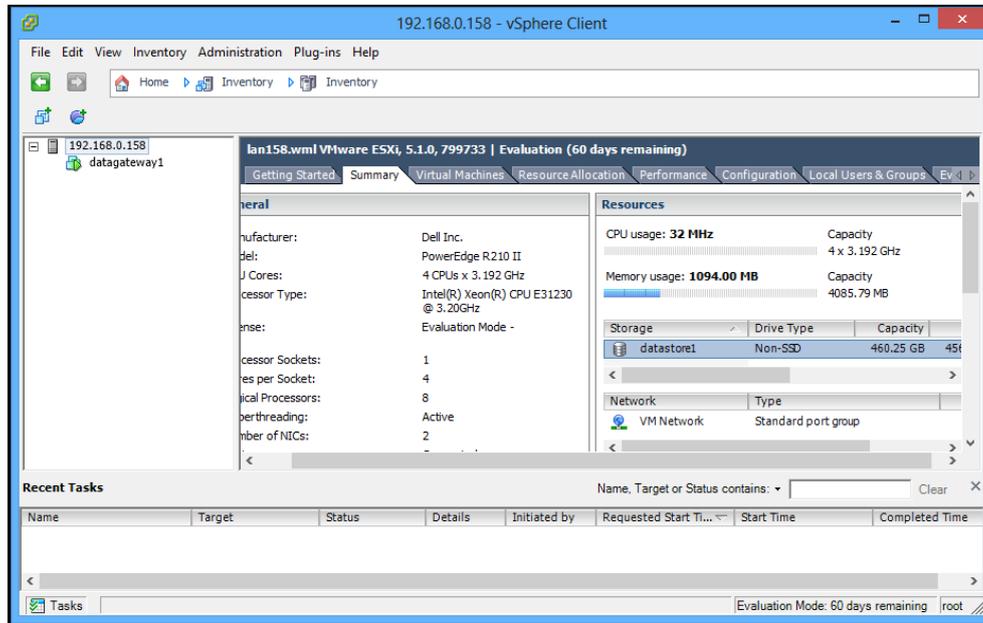
VMware vSphere ESXi walk-through

This walk-through demonstrates the set up of a Virtualized appliance on a VMware ESXi hypervisor version 5.1.0. The ESXi hypervisor is controlled from vSphere client 5.1.0 running on Windows 8.

Note

It is assumed that you have a server with VMware ESXi version 5.1.0 already installed on a server, and VMware vSphere client version 5.1.0 on a separate PC or laptop. For help installing VMware vSphere, please refer to the VMware website.

Figure 2.30. The Summary tab in vSphere client for the ESXi host (192.168.0.158).



Creating a new Virtualized appliance is carried out in the VMware vSphere client user interface. This is accessed via an icon in your start menu or on the desktop. When you open vSphere client, you must connect to the IP address of your VMware ESXi host.

Once you have connected to the ESXi host, you will see the vSphere client **Getting Started** screen. Figure 2.30 shows the vSphere client, with the **Summary** tab showing.

Figure 2.31. VMware images on the WML web server.

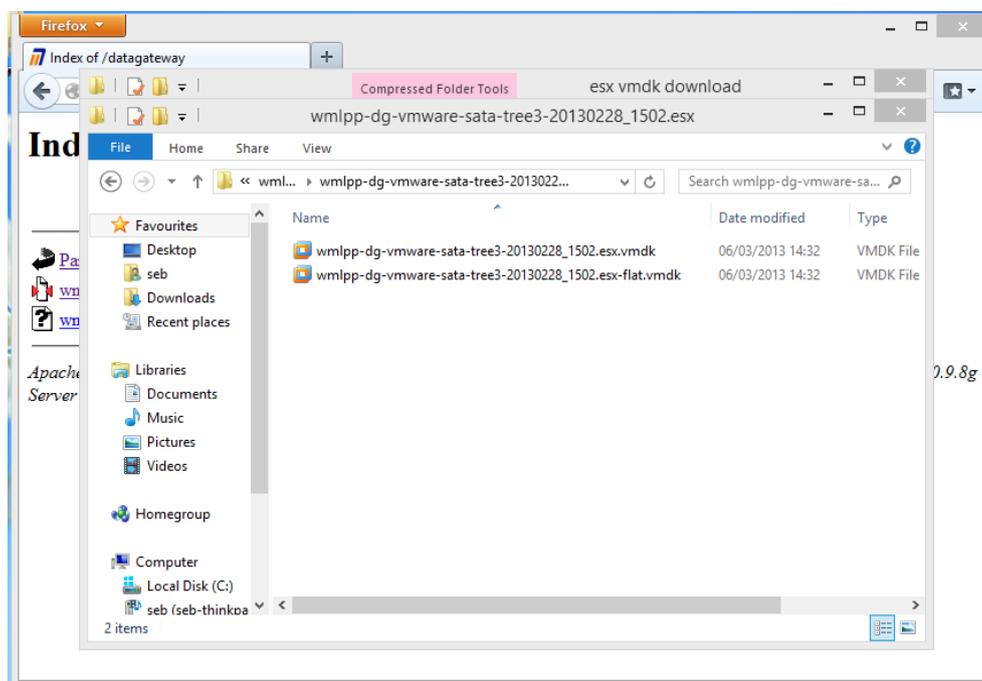


Before starting, download the latest DataGateway image from:

<http://wmltd.co.uk/datagateway/>

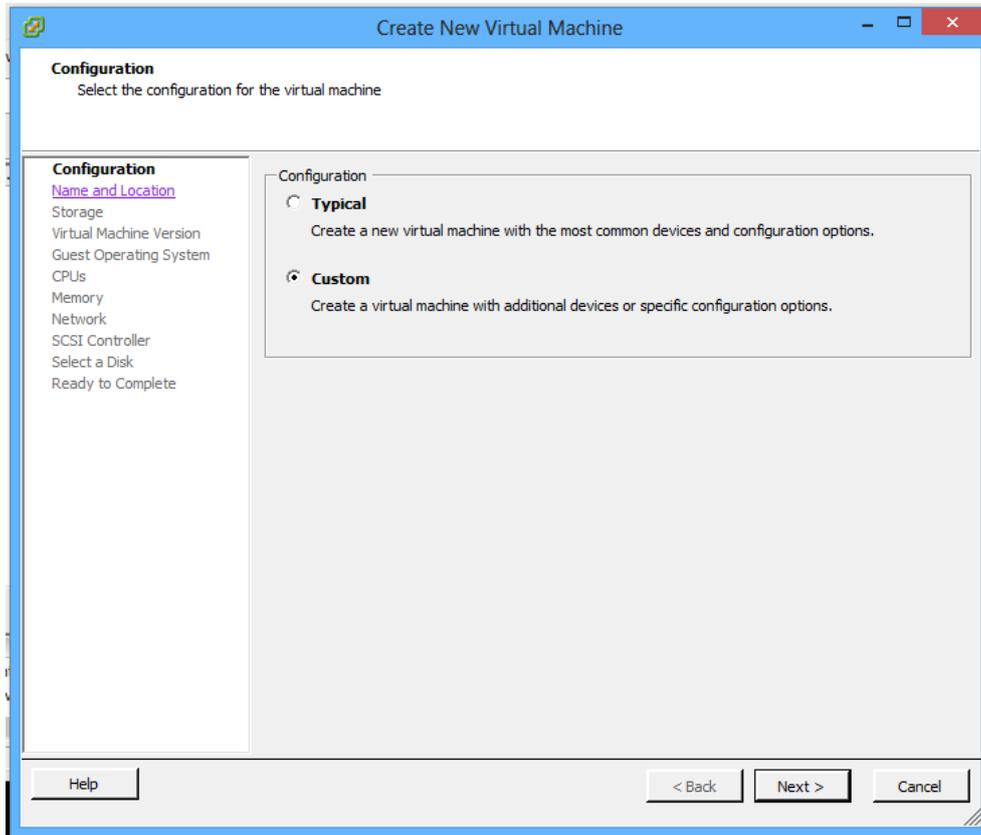
Choose the .esx.zip file (Figure 2.31). Unpack the zip and find the two .vmdk files (Figure 2.32).

Figure 2.32. The .esx.zip file unpacks to provide two vmdk files.



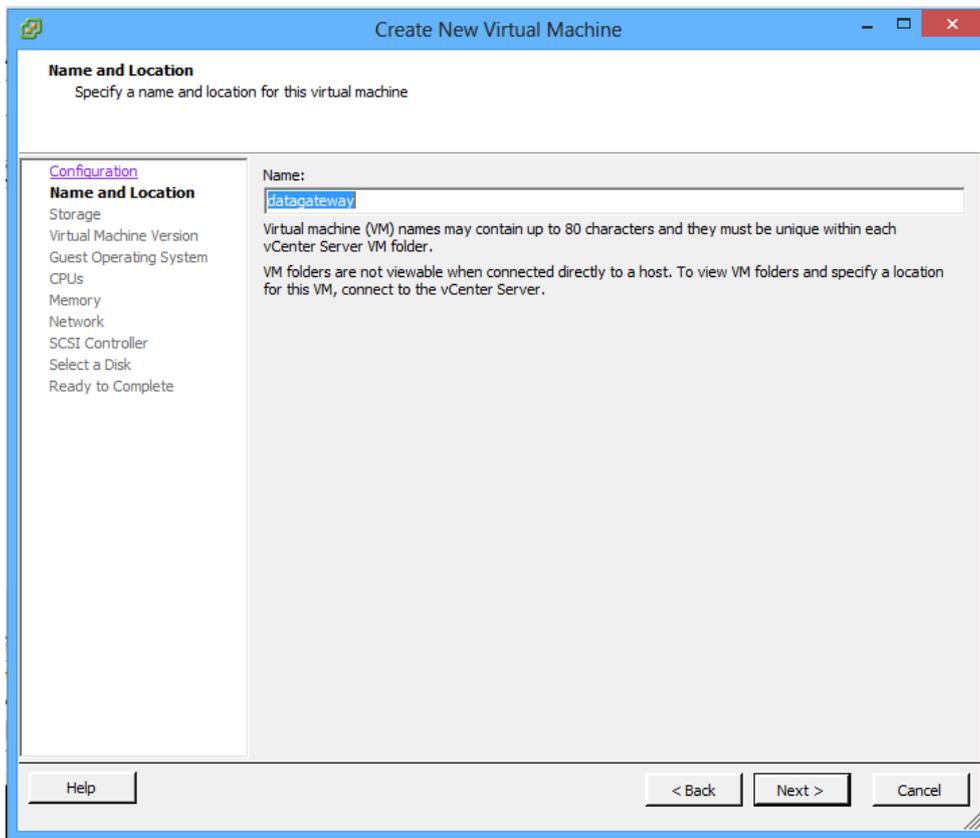
Upload both of the .vmdk files to the datastore of your ESXi host; in the **Summary** tab (Figure 2.38), right-click on the datastore and choose **Browse datastore**. The Datastore Browser has icons to upload two files. Your datastore will now contain the pre-installed virtual disk for your new virtual machine.

Figure 2.33. The “New Virtual Machine” wizard.



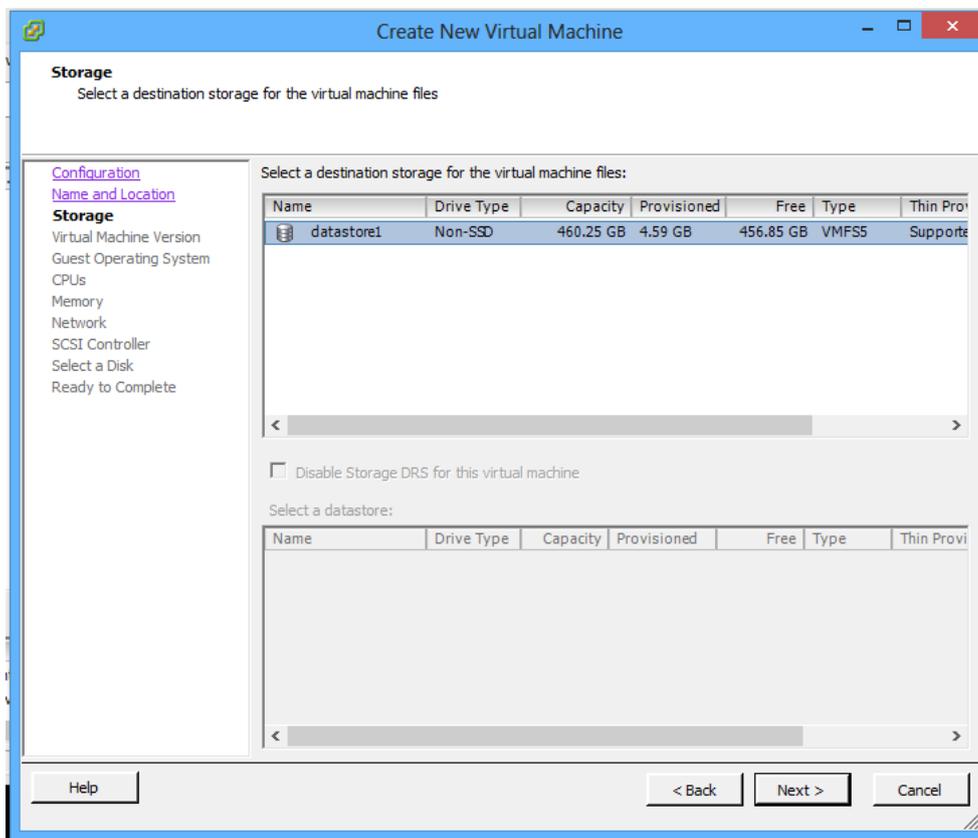
In the **Getting Started** tab, click the link **Create a new virtual machine**. The **Create New Virtual Machine** wizard will open (Figure 2.33). Select **Custom** so that it will be possible to specify that we want to use an existing, preformatted virtual disk. Click **Next**.

Figure 2.34. Choosing the name for the virtual machine.



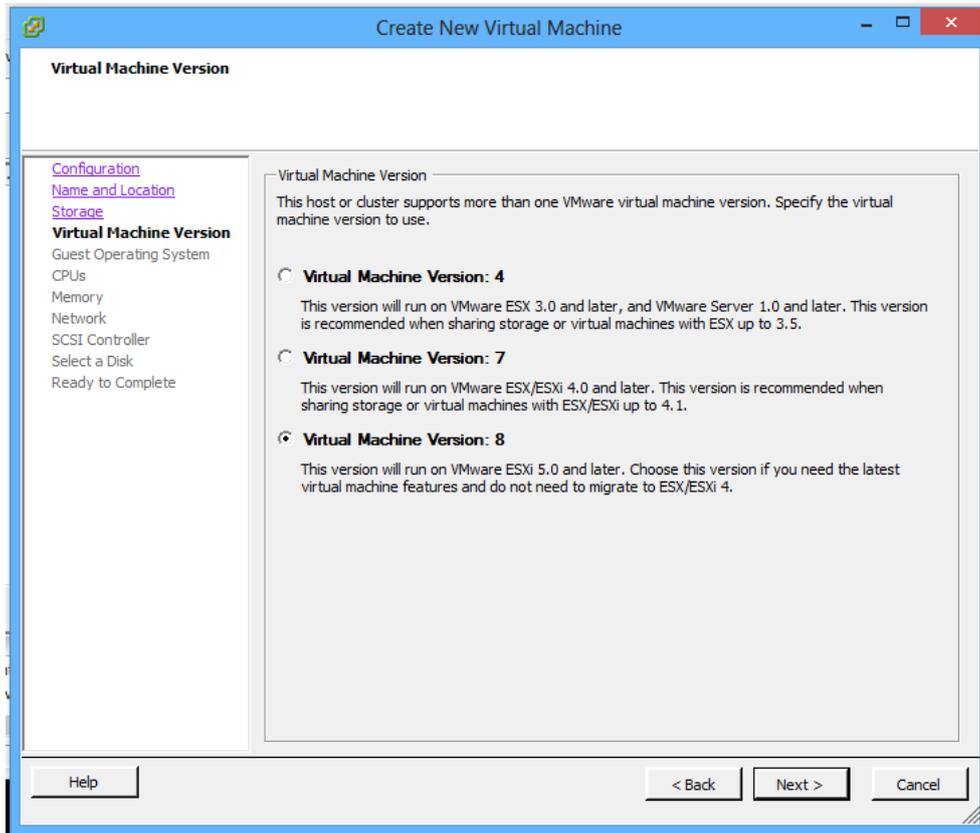
You can now choose the name for your virtual machine. Here, it has been named “data-gateway” (Figure 2.34).

Figure 2.35. Choosing a datastore for the virtual machine.



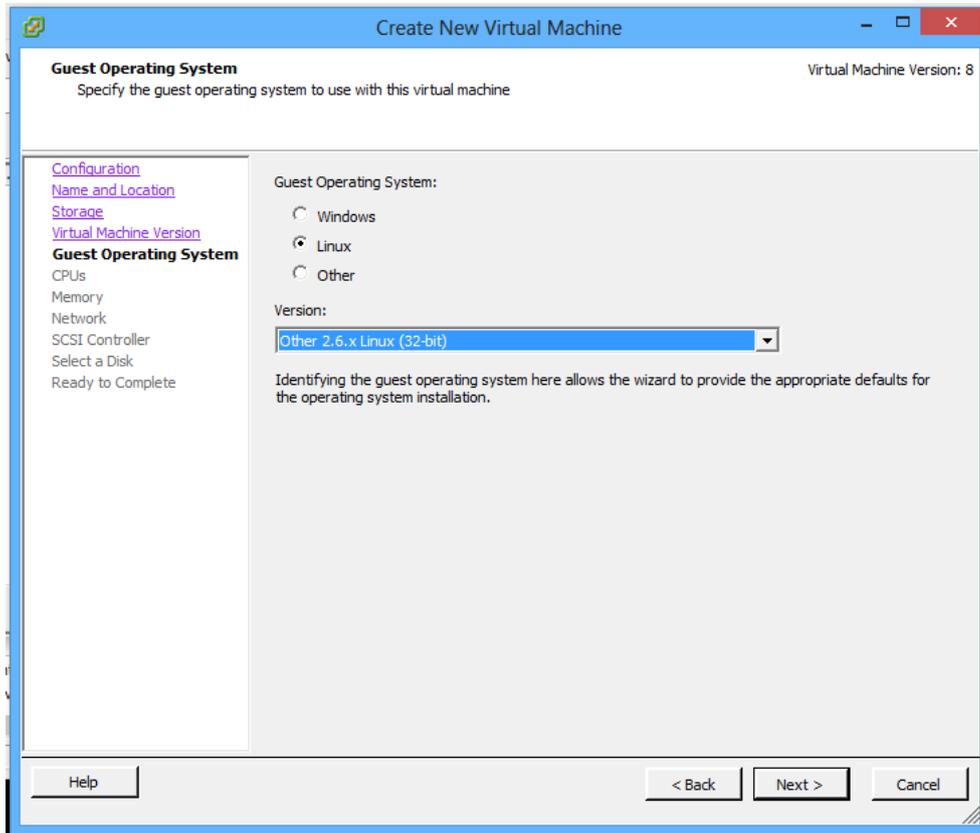
The next window allows you to specify a datastore in which you will create or find the virtual disk for the machine. Choose the datastore to which you uploaded the DataGateway vmdk files. In this example, there is only a single datastore called datastore1.

Figure 2.36. Choosing the virtual machine hardware compatibility.



The Virtual Machine Version window allows you to specify the hardware compatibility for the virtual machine. This may be a consideration if you wish to move your virtual machine to older versions of VMware ESXi or to other VMware infrastructure products. In our example, we leave it at its default of **8**.

Figure 2.37. Specifying the operating system for the guest appliance.



The next window (Figure 2.37) allows you to choose which operating system your target appliance will be running. All WML appliances run a 32 bit Linux operating system and so you should choose **Linux** for **Guest Operating System** and **Other 2.6.x Linux (32-bit)** for **Version**. Click **Next**.

Figure 2.38. Setting the number of processors for the virtual appliance.

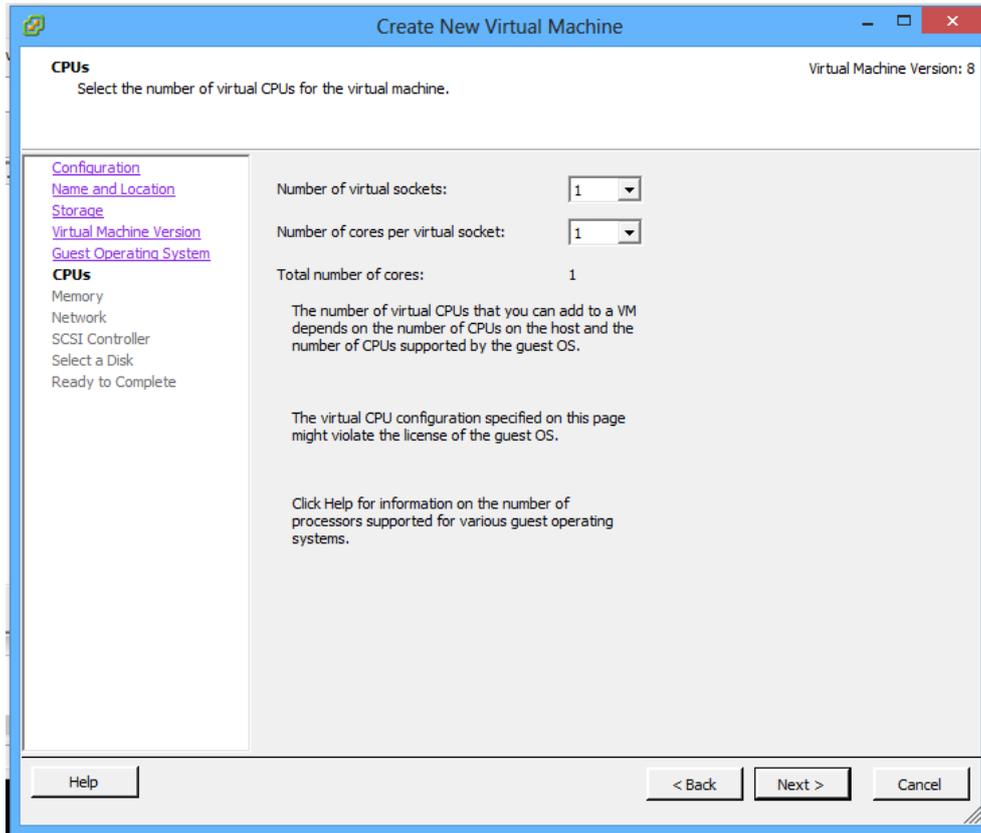
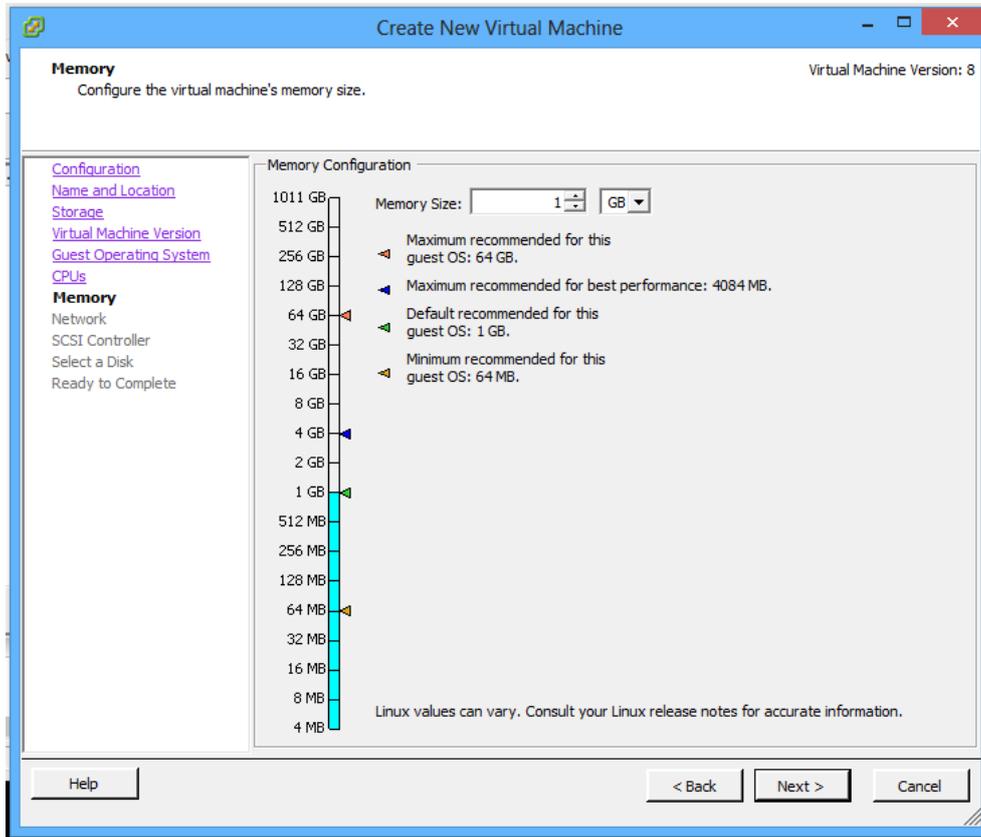


Figure 2.39. Setting the amount of RAM for the virtual appliance.

The next two screens (Figure 2.38 and Figure 2.39) allow you to specify the initial amount of RAM and processor resource to assign to your virtual print appliance.

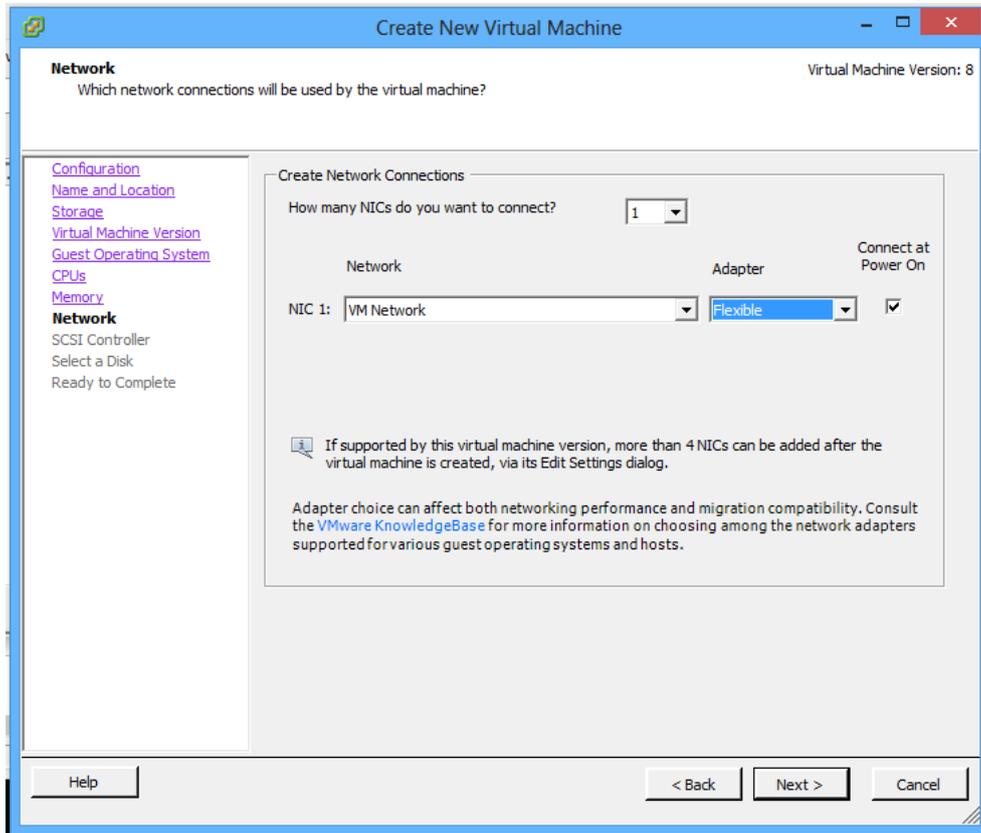
Assign the desired number of processors for the virtual appliance. A single processor may be sufficient for simpler applications. Assigning two processors to the system will improve the responsiveness of the web interface when the print processing load is high.

Specify a minimum of 256 MB of RAM. Some applications may require more memory to run correctly, but 256 MB is sufficient for simpler tasks and short documents.

Note

The amount of RAM may be altered after the virtual appliance has been created. Altering the number of processors is not recommended by VMware, but it can be done. WML appliances will safely handle a change in the number of available processors.

Figure 2.40. The Network window.



In the **Network** window (Figure 2.40) you can select the number of virtual network interface cards to assign to the virtual machine. The default of a single NIC is usually suitable for a DataGateway. By default, the NIC for the virtual machine will join the same network as the ESXi host.

Figure 2.41. Selecting the virtual I/O controller type.

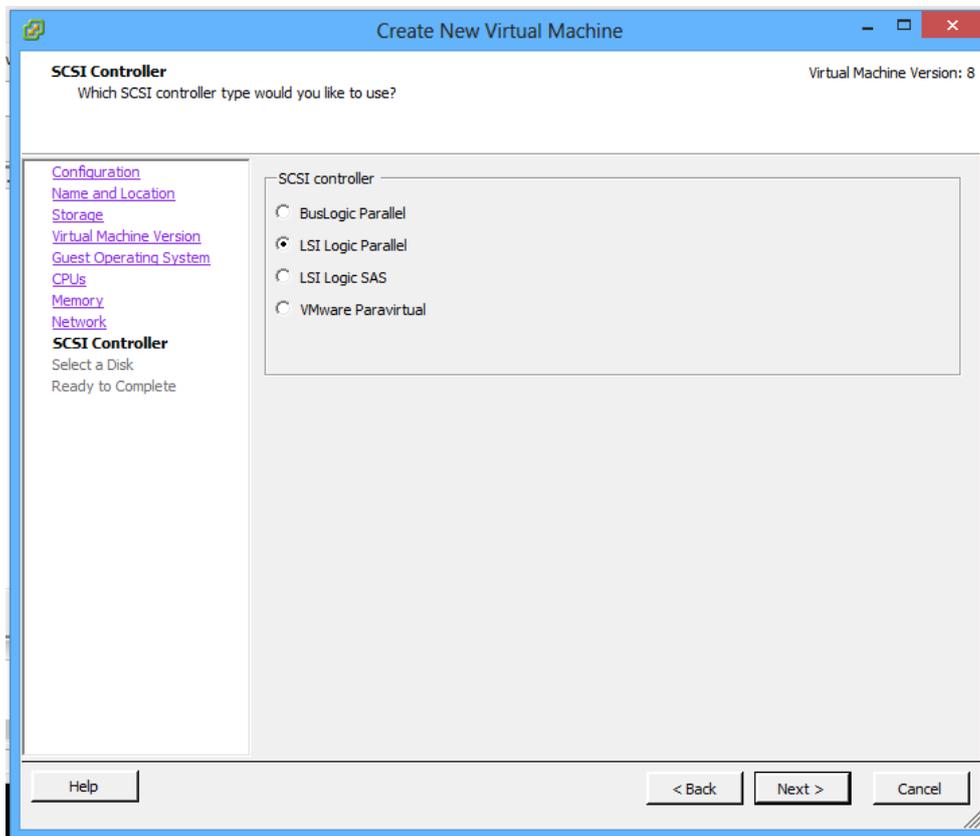
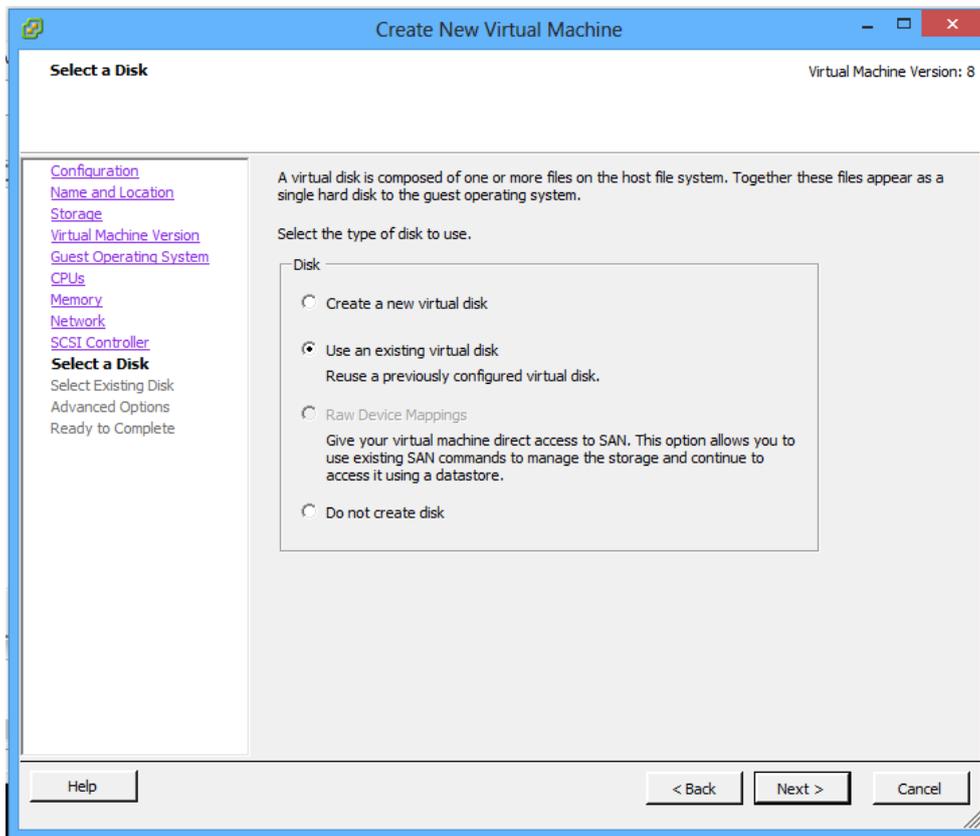


Figure 2.41 shows a screen allowing a choice of virtual SCSI disk controllers. Leave the default **LSI Logic Parallel** checked.

Figure 2.42. Selecting a hard disk for the virtual appliance.



In the next window (Figure 2.43) you have the option to **Create a new virtual disk**. This would create an *empty* disk, which would then need to have the operating system installed (perhaps from a dvd or cdrom). Because you are using a pre-populated disk file supplied by WML, you should click on **Use an existing virtual disk**, which will open the next window, shown in Figure 2.43.

Figure 2.43. Selecting an existing disk.

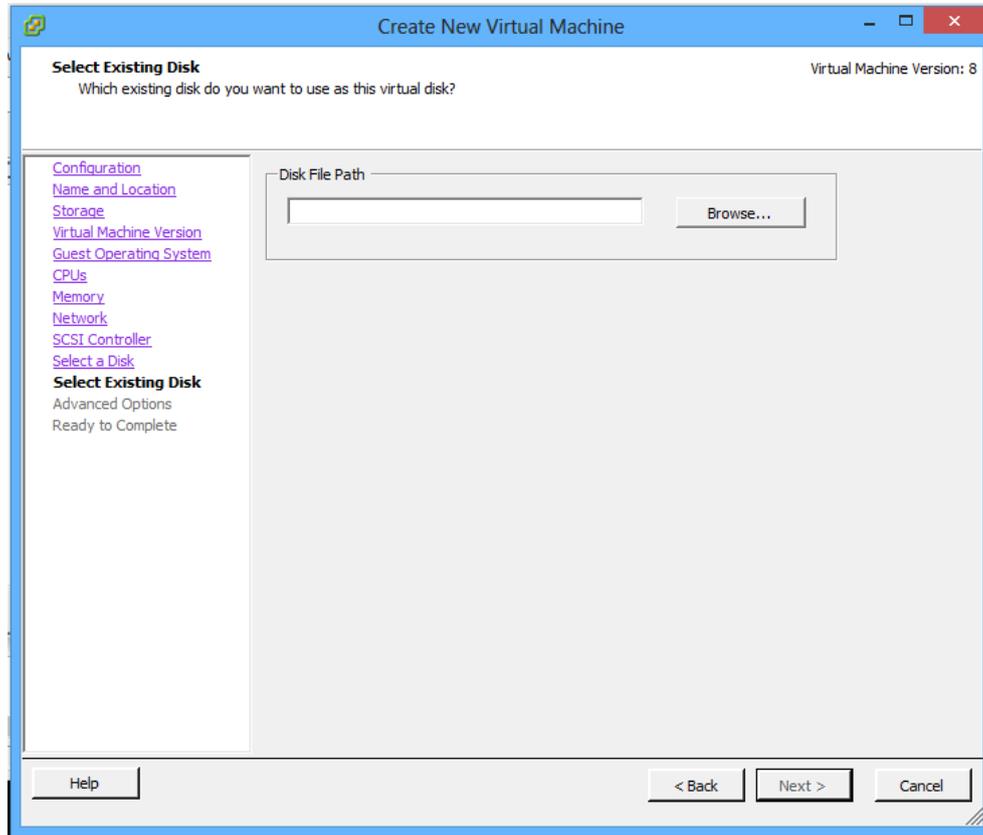
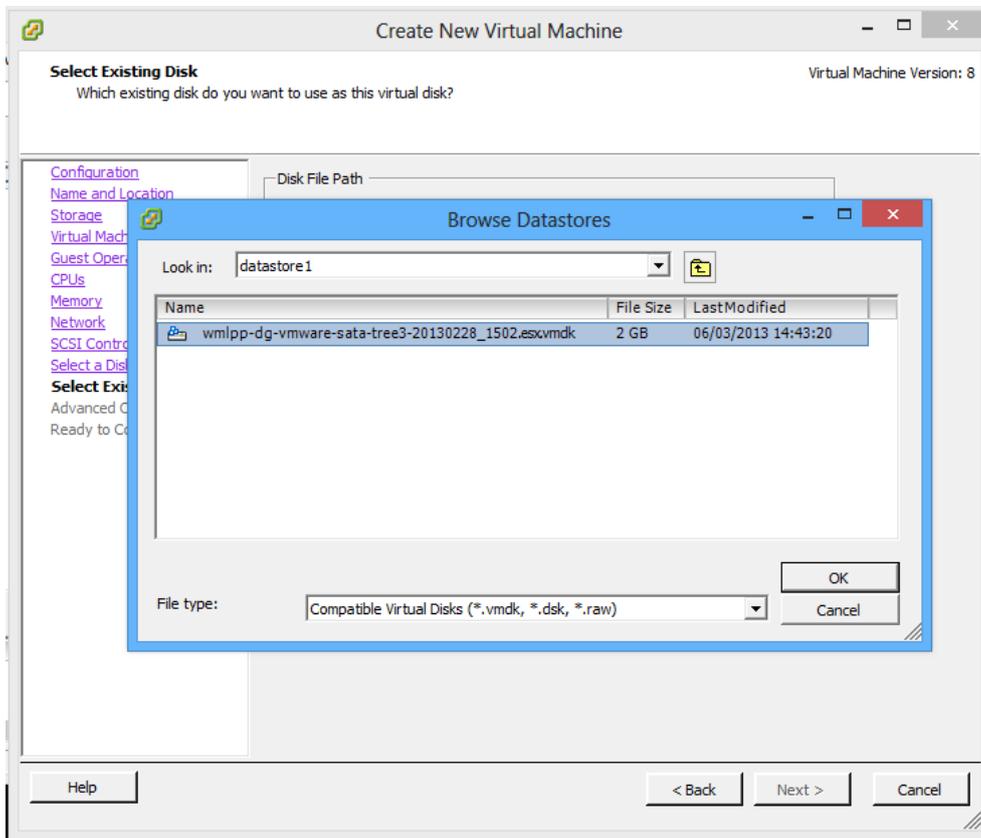
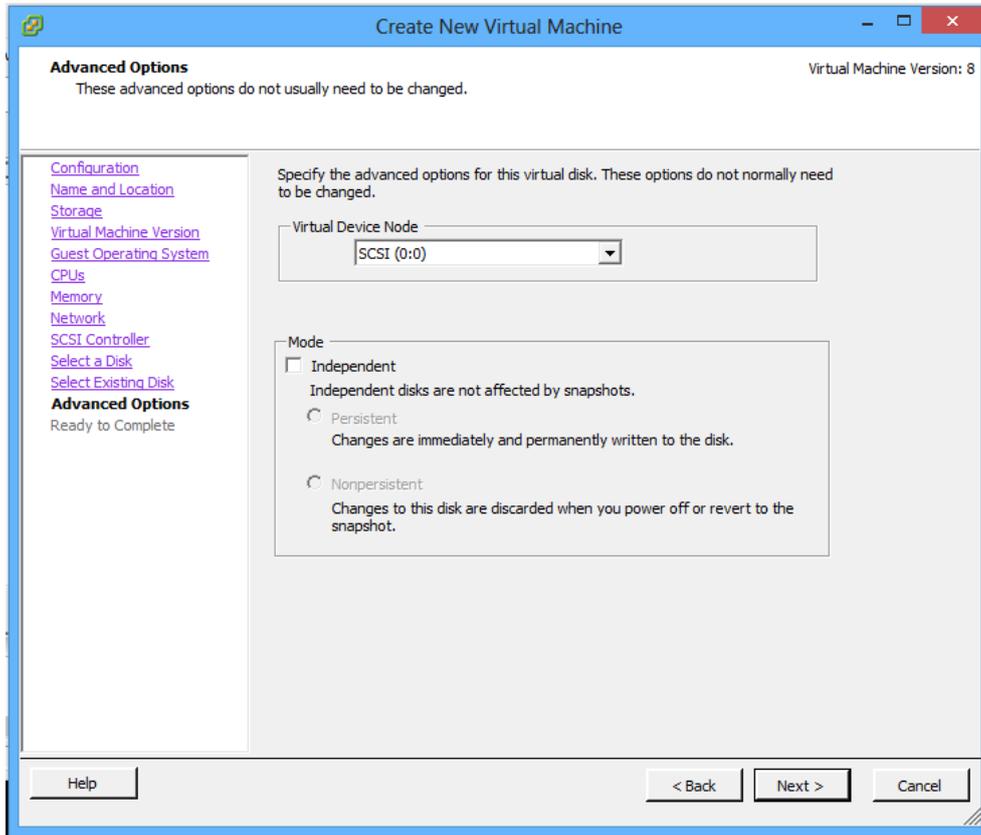


Figure 2.44. Selecting the WML Virtualized disk file.

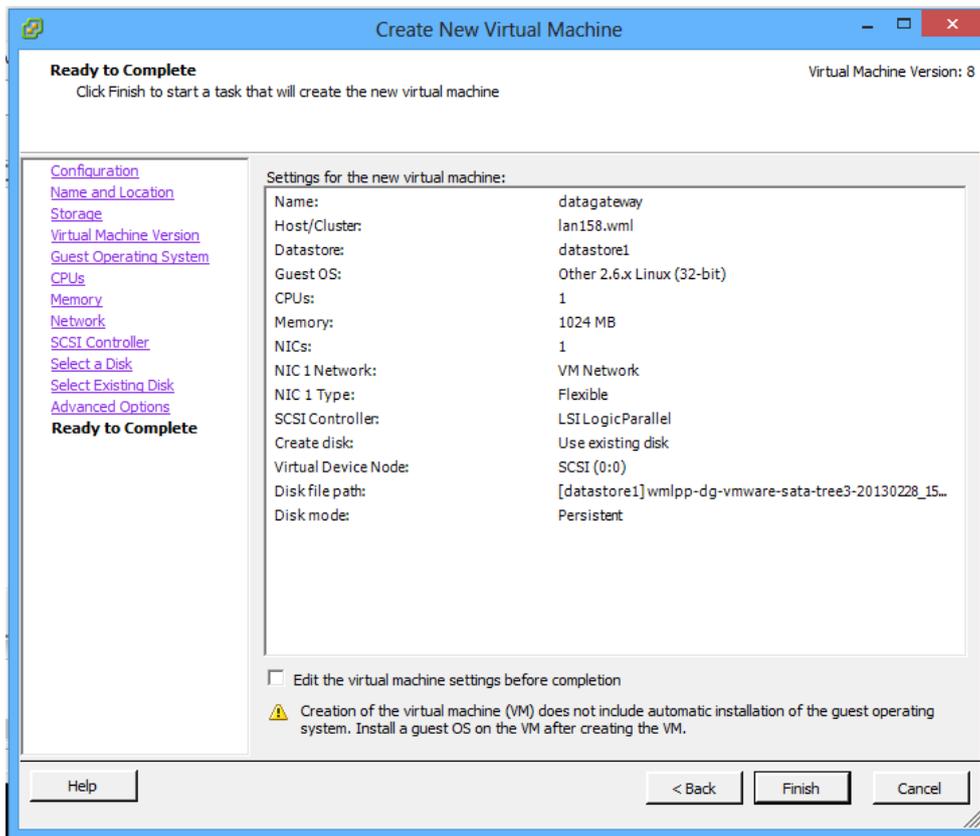


Browse to the vmdk disk image (Figure 2.44) in your datastore and select it.

Figure 2.45. Advanced disk options.



The penultimate window shows some advanced options for your virtual disk. Leave the virtual device node as **SCSI (0:0)** and click **Next**.

Figure 2.46. Ready to Complete - a summary of your new virtual appliance.

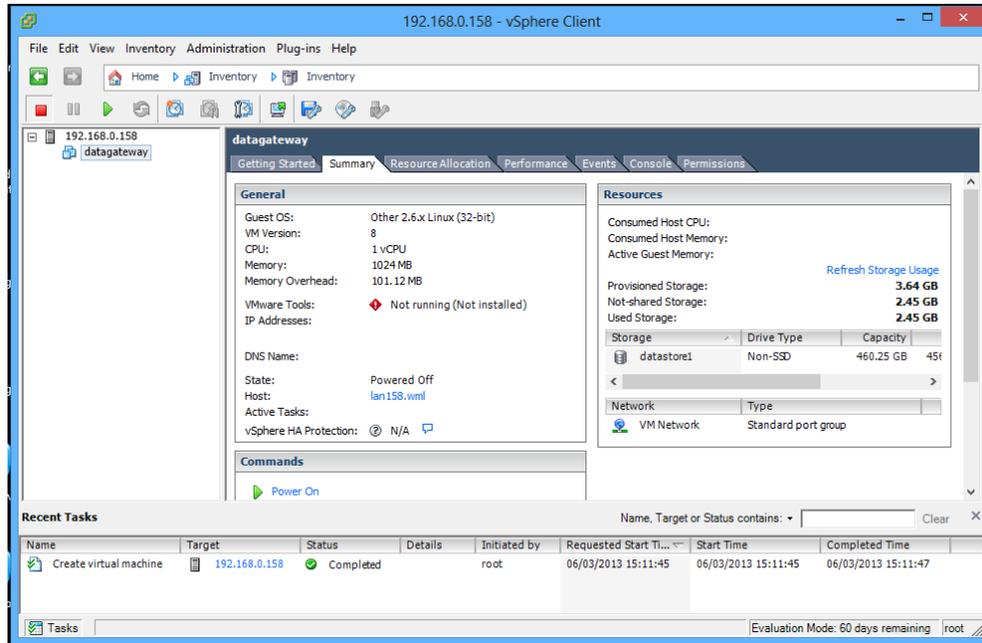
The last window shows a summary of the settings chosen for your new virtual machine (Figure 2.46). Click **Finish** to create the VM.

After creating your new virtual machine, a new item will appear in the inventory for your ESXi host (Figure 2.47). You can also access the console of your virtual machine in the vSphere client to find its IP address. Connect a browser to the IP address to access the DataGateway web user interface.

Note

There is no need to install VMware tools on a WML Virtualized. VMware tools enhances the user experience when using a mouse within a graphical environment. The DataGateway has no graphical console and hence VMware tools is not used.

Figure 2.47. The new appliance summary tab.



We hope you found this walk-through for VMware vSphere and ESXi useful.

2.2. Accessing the console

Occasionally, you may need to access the console of your DataGateway, to carry out system configuration, maintenance and trouble-shooting.

Accessing the console on Tempest appliances

You will need a USB keyboard, a VGA monitor and an appropriate VGA cable to access the console of Tempest and Server-based appliances.

Connect a VGA monitor and USB keyboard to the DataGateway. When the DataGateway is switched on and the system has booted up you will see a login prompt.

You can log in as **admin** for manual IP address configuration or as **root** for other system tasks.

Accessing the console on Virtualized appliances

The virtual machine console for Virtualized appliances is available via the VMware Server management interface, the VMware Workstation UI or the vSphere client, depending on which VMware infrastructure is in use.

You can log in as **admin** for manual IP address configuration or as **root** for other system tasks.

Accessing the console on Server-based appliances

Please contact WML support for details of how to access the console on your Server-based DataGateway.

2.3. IP address configuration

By default, the IP address of the DataGateway will be allocated using DHCP.

To determine what IP address has been allocated you can consult your DHCP server logs. Alternatively, you can log into the DataGateway console, view the IP address displayed there or carry out manual IP address configuration.

Console IP address configuration

When you need to configure the IP address without the web user interface, you can do so by logging in at the machine's console, as described in Section 2.2.

At the DataGateway console, log in using the username **admin** and password **blankdoc** (or the password that has been set for the admin user on your system).

When you have logged in, the current network settings will be displayed. You will then be given the option to configure the IP address manually or via DHCP, or to quit.

```
Enter "manual" for manual IP configuration
or "dhcp" for automatic IP configuration (q to
quit).
```

If manual configuration is chosen, you will then be prompted for an IP address for the device, a Netmask address and, optionally, a gateway address.

```
Enter the IP for the device: 192.168.0.20
Enter the Netmask for the device: 255.255.255.0
Enter the Gateway (n for none): _
```

For both manual and DHCP configuration, you will be asked to confirm whether you want to reset the DNS settings. Specifying a DNS server IP which is not accessible on your network can lock the web based user interface of the DataGateway.

```
Do you wish to reset the DNS settings? Specifying a DNS
server IP which is not accessible on your network
can lock the web based user interface of this device.
Enter 'y' to remove DNS settings, 'n' to keep them.
```

The DataGateway will now be reconfigured, the new network settings will be displayed and the program will exit. To make further changes you will need to log in and follow the process through again.

2.4. Accessing the DataGateway user interface

You can access the DataGateway user interface from anywhere on your local network by directing a web browser at the IP address that has been allocated, either manually or via DHCP, to the device.

You will need to log in to administer and configure the DataGateway.

DataGateway users

The DataGateway is configured with two levels of user: *System Manager* and *Print Manager*. The default passwords shipped with the system are “blankdoc” in each case.

The System Manager has unrestricted access to all configuration options on the DataGateway.

The Print Manager has restricted access to the system, allowing the user to administer print jobs and print queues, but not make changes to system settings or operate the optional LAN Manager. For example, the Print Manager can't change the IP address of the DataGateway or re-direct system logging messages.

Chapter 3. System Maintenance

This chapter covers information relevant to the day to day maintenance of the DataGateway.

As the DataGateway was developed as an appliance form-factor system, with a custom-compiled operating system, there are relatively few regular tasks to be carried out by systems admins. There are a number of resources which should be monitored to ensure that production is uninterrupted. In most cases, alert emails are sent out to systems admins as a resource (such as local disk storage) is running low.

3.1. Local storage

The DataGateway provides local storage for log files, archive files and user-supplied PDF forms. When the available storage drops below 20%, “persistent filesystem alerts” are sent out to the main email alert list (as configured in **System Settings > Email**).

To clear unwanted files and free up space, connect to the system using **SCP** or **SSH**. Users of Microsoft Windows can use **WinSCP** for file access using the SCP protocol. SSH is implemented by **putty** on Windows.

The following directories/folders will contain dated files. Old or unwanted files can be archived off or deleted as necessary.

Warning

Do not remove files from any locations other than those listed here unless specifically instructed to do so by your vendor. Removal of any other files may affect the correct operation of your DataGateway.

- **/etc/wml/log/** contains system log files and report files. The log files are named with the formats **YYYY-MM-DD-wmlpp_job.log** and **YYYY-MM-DD-wmlpp_sys.log**. The CSV reports which are generated from the logs have the format **YYYY-MM-wmlpp_report.csv** or **YYYY-MM-DD-wmlpp_report.csv**. Outbatch logs may also be present, which have the format **YYYY-MM-DD_hh-mm-ss-QUEUE_NAME_ob.log**.
- **/etc/wml/archive/** contains archived print files. The format of these files will depend on the user-specified file naming for each queue (as set in **Print Platform > Queue Config > Archive**). Each queue will have a sub-directory within **/etc/wml/archive/**.
- **/etc/wml/cdlarchive/** (applicable only to systems which are licenced with the CDL preprocessing module) contains archived data as received from the CDL software, before any processing by the DataGateway. A sub-directory for each CDL-processing (or analysing) queue will be present containing files with the format **YYYYMMDD_hhmmss.raw**
- **/etc/wml/master/** (applicable only to systems which are licenced with the Appliance Manager module) contains firmware upgrade images which have been distributed to managed appliances. These have the format **wmlpp-PRODUCT-kvm-sa-**

ta-treeX-YYYYMMDD_hhmm.img where PRODUCT will be dg for DataGateway, pre for PrePrint and pplus for PdfPlus.

- **/etc/wml/outbatchshare/** (applicable only to systems which are licenced with the proprint module) contains completed “output batch” zip files. These have a user-definable format. Generally these would be stored on a network share, but it is possible to have them stored on local storage.

Note

Do not be confused by the existence of **/etc/wml/outbatch**. This contains in-progress print files which will either be released to a printer, or combined into a completed output batch zip file. Both **/etc/wml/outbatch** and **/etc/wml/outbatchshare** may exist in normal operation.

Chapter 4. Disaster recovery and backup procedures

This chapter describes strategies for making a backup of your production DataGateway so that a replacement system can be re-created in the event of it failing.

Warning

Please note that WML's standard maintenance contracts do NOT cover the re-build of a solution following a hardware failure unless virtual machine backup images can be provided or a WML disaster recovery system is in place.

4.1. Backup

Automated backup with WML's disaster recovery system

The best way to backup your DataGateway is to implement WML's disaster recovery system. This is implemented as a feature in the Appliance Manager module (see Section 12.6).

The production system is configured as the “DR master” and has the Appliance Manager module installed. A second, backup DataGateway is configured to act as the “DR backup”. The DR master automatically pushes all configuration and archive files to the DR backup twice per day. It also makes a database dump of the mysql database and pushes this to the DR backup.

Crucially, every time a firmware upgrade is applied to the DR master, the upgrade is pushed to the DR backup, which automatically applies the upgrade itself. This means that the DR backup can be rebooted into production mode at any time, with the on-board configuration being no more than 12 hours out of date.

Warning

The DR backup system should be running on the same infrastructure as the main system. Either both should be KVM (Kernel Virtual Machine) based (such as those which are supplied by WML as hardware appliances) OR both should be VMware based. If one system is KVM based, and the second is VMware based, firmware upgrades applied to the main system will not propagate to the DR backup system.

Static virtual machine backup procedure

If a WML disaster recover system is not implemented, it is important to make backups of the virtual machine disk image files.

Note

All Server-based and Virtualized DataGateways operate as virtual machines. Server-based systems are provided running on a pre-configured Ubuntu Linux server which provides a Linux KVM infrastructure for running the DataGateway virtual machines.

Linux KVM virtual machines

If you have a Server-based DataGateway or you have implemented DataGateway on your own KVM infrastructure, you will be running a KVM based system.

If you are running DataGateway on your own KVM infrastructure, please make regular backups of the DataGateway qcow2 disk files, and of any files stored in a logical volume which is attached to the DataGateway virtual machine.

Note

It is not necessary to shut down the DataGateway while copying the qcow2 disk files, but it is advisable. If the DataGateway is not shutdown when the backup is made, it will be necessary to run a file system check on the qcow2 file when the recovered DataGateway boots up. Although this is completed automatically on every DataGateway boot, there may be rare cases where the filesystem check cannot recover a file in the disk image.

If your KVM infrastructure has the facility to make complete backups or copies of running virtual machines, then consider making use of it.

If you are running a Server-based DataGateway, you will need to backup the qcow2 disk files in **/var/vm/** on the Ubuntu server and any logical volume which has been presented to your DataGateway as a second virtual disk.

To backup files in **/var/vm/**, use the scp (Secure CoPy) protocol to connect to the server on port 22 or 1522. Copy all files from the path **/var/vm/** on the remote system to your local backup location. You will need to have obtained and recorded the password associated with the user **wmldev** on the Ubuntu server.

Note

The WinSCP program (<http://winscp.net/>) is a free utility which implements the scp protocol and runs on Microsoft Windows.

You can determine if your DataGateway is using an attached logical disk by consulting **System Settings > Status/Control > Share mounts**. This will open a dialog containing a table of **File Systems and Mounted Network Shares**. If you have a **Path** of **/dev/sdb** with **Mount Point /etc/persistent**, then you do have an attached logical volume.

Because this logical volume may be encrypted, the best way to back it up is to copy the files from the DataGateway rather than via the underlying server.

Use scp to copy all files from `/etc/persistent/` on the DataGateway to your local backup location. Use port 22 or 7522.

VMware virtual machines

If you are running your DataGateway on a VMware infrastructure, please make use of the tools provided by VMware to make backups of your DataGateway.

Database backups

An on-board MySQL database is used to record information for the Document Content Control module, the Electronic Document Delivery module and for full audit-trail reporting.

A static virtual machine backup will copy the existing state of the database, but it is also possible to make external backups of the database content.

To make dumps of the on-board database, ask your vendor to set up a read-only database user and then make use of the mysqldump utility to extract a snapshot of the database contents from the DataGateway.

Making backups of DataGateway system settings

The DataGateway web user interface provides a method for saving configuration files, such as queue configuration and Document Content Control configuration.

These backups are particularly useful for creating duplicate systems for testing or to provide to WML to aid support. Each backup created is a tar archive file, containing the backed up files.

Refer to the section called “System settings” for details of backing up DataGateway system settings.

Warning

This method does NOT allow you to backup archived documents, database content (used for Document Content Control, Electronic Document Delivery and audit report systems) or Archive Manager configuration.

4.2. Recovery

The method for recovering a DataGateway following a failure of the hardware platform will depend on how the backups were made.

If the WML disaster recovery system is implemented, the DR master should be unplugged from the network, if it is still running, and the DR backup should be rebooted into “master mode”. (see Section 12.6 for more details).

If you are managing the virtualization infrastructure, then follow suitable steps to re-create a virtual machine from the backups you have made of the disk images/virtual machine.

If you are running a Server-based DataGateway it will be necessary to have WML supply a preconfigured server running Ubuntu Linux with KVM onto which the virtual machine backups can be restored.

Disaster recovery and backup procedures

Note

The standard WML software maintenance contract will cover the work required to restore virtual machine disk image files onto the server. Hardware maintenance agreements should be arranged to cover the server hardware, which is otherwise supplied at additional cost.

If the system is a Tempest appliance, a replacement should be set up using the backup files made in the DataGateway web user interface.

Chapter 5. DataGateway Configuration

This chapter covers the steps necessary to configure the DataGateway system. After configuration, the system will have the correct network settings, it will keep time correctly and system log messages will be logged to your syslog server.

After configuring the system, you will be ready to set up print queues, as described in Chapter 6: *Queue Management*.

This chapter also describes the configuration of passwords for authentication and describes how to modify the licence key.

5.1. System configuration

System configuration tools can be found in the DataGateway web user interface by selecting the **System Settings** item from the main menu.

Network settings

To access the network settings controls (Figure 5.1), select **System Settings > Network**. This notation means “select the System button from the upper row, then select the Network tab in the lower row.”

Figure 5.1. Network settings

Print Platform | File Manager | **System Settings** | Password | Settings | Firmware | Status/Control

Network | Time | Location | System Log | Email | Licences

Network Settings

Network address for eth0

Use dhcp (numbers below will be ignored and you will need to manually re-direct your browser to the new IP address)

Hostname:

IP Address:

Netmask:

Gateway:

DNS settings

Use DNS (if unchecked, numbers below will be ignored)

Primary DNS server:

Secondary DNS server:

DNS Domain:

DNS Search Order:

Output bandwidth limit

Set the bandwidth limit (for sending data). Specify limit in bits per second or use K, M or G. Remove the limit by entering "0"

Bandwidth (bits):

Here you can set the network address for the network interface(s). Figure 5.1 shows a Pdf-Plus on which there is one network interface, **eth0**. On hardware platforms with multiple

network interfaces, additional sections for the extra interfaces will be shown. The interfaces are named **eth0**, **eth1**, **eth2** and so on.

The parameters for each network interface can be set to fixed, static values or can be set to be dynamically updated using DHCP (Dynamic Host Configuration Protocol). If you check “Use dhcp”, the IP address will be allocated by your DHCP server. The client ID passed to the DHCP server will be “eth0” for the **eth0** interface, “eth1” for the **eth1** interface and so on. You may need to consult the log of your DHCP server to find the IP address allocated to the DataGateway. Alternatively, access the console of the DataGateway and login as admin, as described in Section 2.2 in Chapter 2: *Getting Started*. On doing so, the current network settings will be displayed.

Configuring multiple aliases

The DataGateway can be configured with multiple IP addresses for each network interface. To add an alias, press the “Add Alias” button. A field will appear in which you can specify a static IP address or enter **dhcp** to have the alias obtain a dynamic IP address from your DHCP server.

When using DHCP to allocate IP addresses to multiple aliases on the same network interface, you will need to configure the server to make use of the *client ID* parameter. The client ID is an identifier passed by the DHCP client to the DHCP server to identify the request. The first alias on the interface **eth0** has the client ID **eth0:1**, the second alias has the client ID **eth0:2** and so on.

Configuring DNS settings

DNS settings are also configured on the Network Settings page. You can configure the primary and secondary DNS servers, along with the domain and a domain search order for name resolution. These parameters are placed in a standard Unix/Linux **resolv.conf** file.

Bandwidth limit

It is possible to limit the rate at which the DataGateway sends data on its output. This feature can be used to preserve a portion of a site's bandwidth for other uses which may be sensitive to latency.

Note

It is not possible to limit the rate at which data comes *into* the appliance. This is a fundamental limitation of the IP protocol, which is designed to accept all incoming packets on an interface. Bandwidth limiting is always achieved by limiting the rate at which data packets are sent.

To limit the bandwidth for a given interface, enter a number in the corresponding “Bandwidth (bits)” field and press “Update”. A suffix of K for kilobits per second, M for megabits per second or G for gigabits per second can be appended.

To remove the bandwidth limit, enter 0 in the “Bandwidth (bits)” field.

System date and time

To access the system date and time controls (Figure 5.2), select **System Settings > Time**.

Figure 5.2. Setting system date and time

The screenshot shows the 'System Settings' menu with 'Time' selected. The page title is 'System Date and Time'. It displays the current system date and time as 'Fri Jul 30 06:10:22 2010 (Timezone: GMT)'. Below this, there is explanatory text about setting the date manually or using NTP. A 'Manual Date String' input field is present. Further down, there are checkboxes for 'Enable NTP' (unchecked) and 'Use Public NTP Server Pool' (checked). An 'NTP Server' input field is also visible, followed by an 'Update' button.

The system date can be set manually or using network time protocol (NTP). If you choose to use NTP, you can specify to use your own NTP server or a public pool of NTP servers. To make use of the server pool, your device must have access to the public internet and DNS must be configured so that the device can resolve network names.

To set the date, manually enter a string of the form **MMDDhhmmYYYY**. For example, to set date to 09:44 July 19 2006, enter **071909442006**.

To save your changes, click on the “Update” button.

To configure NTP, check “Enable NTP” and select “Use Public NTP Server Pool” or enter the address of your preferred NTP server (you can do both). Note that if your device is unable to

access either the public pool or the preferred NTP server, then a timeout of approximately 10 seconds will occur when you press “Update”.

Don't forget to set the correct timezone (Section : System timezone and locale), to ensure that daylight savings adjustments are automatically applied.

System timezone and locale

To access the system timezone and locale controls (Figure 5.3), select **System Settings > Location**.

Figure 5.3. Setting system timezone and locale

Print Platform File Manager System Settings

Network Time Location System Log Email Licences Password Settings Firmware Status/Control

System Timezone and Locale

Set your location and language/locale here. This will ensure that the system time is always displayed correctly.

Select Timezone: Europe/London (local time is Fri Jul 30 06:14:50 2010)

Select Language: English (en)

Update

The drop-down controls can be used to select your timezone and language/locale.

It is important for time-keeping (see “System date and time”) to set the the correct time-zone in which you are operating.

To save your changes, click on the “Update” button.

System logging

To access the system logging controls (Figure 5.4), select **System Settings > Log**.

The system logging controls allow you to choose whether or not to permanently store system log messages. As well as being useful for diagnosing problems, the system log is essential if you wish to generate print reports (Chapter 11: *Print Reports*).

If left unconfigured, the DataGateway will only store system messages to a temporary log in RAM. This log is frequently truncated and overwritten. In order to permanently store the system log, messages may be sent to a syslog server or to another DataGateway device. If syslog messages are sent from one DataGateway to another DataGateway then they can be encrypted. Messages may also be stored to a network share or local drive (on Server-based and Virtualized systems).

Figure 5.4. System logging

Print Platform File Manager System Settings

Network Time Location System Log Email Licences Password Settings Firmware Status/Control

System Logging [?](#)

View internal system log: [System Settings -> Status/Control](#)

Send syslog messages to: 192.0.0.1 port: 514 protocol: UDP

Receive network syslog messages on port: 514 protocol: UDP

Generate reports from syslog (must store log messages, below)

Email reports (daily)

View full system logs: [File Manager -> Log Files](#)

Store log messages to network share or local hard disk

Share Type: Local Drive

Available SMB/CIFS shared folders: \\WINXP\log Refresh

Manual path entry: _____

You may need to enter your username/password in the fields below to discover the shares on your network.

Share username: _____

Share password: _____

Update

If you want to send syslog messages to another device, check the “Send syslog messages...” box and enter the IP address, port number and protocol to be used. You may send the messages to multiple devices by entering multiple, comma-separated IP addresses.

You may also choose to receive network syslog messages on the DataGateway by checking the “Receive network syslog messages...” box and filling in the port number and protocol appropriately.

To store messages to a network share or local disk, check the “Store log messages to network share...” box and specify the storage type as described in the section called “Selecting network/hard disk storage”.

Note

If you un-check “Send...”, “Receive...” and “Store...”, system log messages will still be written to a file in RAM, but they will be periodically truncated and will not persist after a reboot.

Permanently stored system log messages are stored to two files, `wmlpp_job.log` and `wmlpp_sys.log`. As the file names suggest, `wmlpp_job.log` contains messages relating to print jobs, while `wmlpp_sys.log` stores messages relating to other aspects of the system. Example 5.1 shows an example set of log files.

Example 5.1. Example system log files

```
2009-07-10-wmlpp_job.log           2009-07-12-wmlpp_sys.log
2009-07-10-wmlpp_job.log.partner  2009-07-13-wmlpp_job.log
2009-07-10-wmlpp_report.csv       2009-07-13-wmlpp_report.csv
2009-07-10-wmlpp_sys.log          2009-07-13-wmlpp_sys.log
2009-07-11-wmlpp_job.log          2009-07-14-wmlpp_job.log
2009-07-11-wmlpp_job.log.partner  2009-07-14-wmlpp_job.log.partner
2009-07-11-wmlpp_report.csv       2009-07-14-wmlpp_sys.log
2009-07-11-wmlpp_sys.log          wmlpp_job.log
2009-07-12-wmlpp_job.log          wmlpp_report.csv
2009-07-12-wmlpp_job.log.partner  wmlpp_sys.log
2009-07-12-wmlpp_report.csv
```

wmlpp_job.log and wmlpp_sys.log are “rotated” daily in the early morning. Each day, the files are renamed in the format YYYY-MM-DD-wmlpp_job.log (for wmlpp_job.log) and the current log is then truncated. The system messages for the *previous* day are stored in the dated file. For example, 2009-07-12-wmlpp_job.log contains the job logs for 2009-07-11.

When network share or hard disk storage is available, reports can be generated from the syslog data. Check “Generate reports from syslog”.

If you wish to receive reports by daily email, check the “Email reports” box and configure email settings for the DataGateway as described in the section called “Email settings”. More information about the content of the reports is available in Chapter 11.

Note

The .partner files are present in Example 5.1 because this DataGateway is generating reports as an internet-linked pair of appliances. The .partner logs are pulled from the transfer partner appliance in a case where the system log messages from the partner are required to complete the report.

To save your changes to the system logging settings, click on the “Update” button.

Selecting network/hard disk storage

The “Network/Hard Disk Storage” section enables you to specify a network share or local drive (on Server-based systems) to use for system log storage. The network share may be either a Unix/Linux Network File Server (NFS) share or a Windows/Samba (SMB/CIFS) network share. The preferred storage option may be selected using the “Share Type” drop-down.

Selecting a share

NFS (Linux/Unix shares)

To connect to a Linux/Unix NFS share, choose **NFS (Linux/Unix)** from the **Share Type** drop-down. Enter the path to the NFS share in the following format:

[Network/IP Address]:[folder path]

For example:

192.168.100.100:/input_data

The NFS server's exports file should be configured to allow both read and write access to the directory path (/input_data in the example). An example of a suitable line in /etc/exports on the NFS server is:

```
/input_data 192.168.0.0/255.255.255.0(rw,no_root_squash, sync, subtree_check)
```

If the NFS share is to be used as an *archive* Share, then correct permissions must be set on the share directory on the NFS server. The following would be required for an archive share specified as:

192.168.100.100:/archiveData

On the server it would be necessary to change the user on this directory to “root” or 0 and the group of the directory to 7, which is the GID for the “lp” user on some systems:

```
root@nfsServer# chown 0:7 /archiveData
```

Finally, ensure that both user and group are able to read and write to the directory:

```
root@nfsServer# chmod ug+rwx /archiveData
```

SMB/CIFS (Windows shares)

For Windows SMB/CIFS network shares, select **SMB/CIFS (Windows share)** and click on the **Refresh** button to scan the network for available shares. This will populate the **Available SMB/CIFS shared folders** drop-down.

Important

Most recent Microsoft Windows file servers will not broadcast their available shares to the network without authentication.

If necessary, enter your network username and password in the fields **Network user name** and **Network password** and then press **Refresh**. The username and password will be used to obtain the list of available shares.

It is also possible to enter the share path manually in the **Manual path entry** text field, by choosing the option **Enter path in text box** from the **Available SMB/CIFS shared folders** drop-down.

Enter the SMB/CIFS path in the format:

```
\\MACHINE\shareName
```

The **Network user name** and the **Network password** fields must be filled in to attach shares which require authentication.

On Server-based systems you may choose to use a local drive for storage.

To save your changes, click on the “Update” button.

Email settings

To access the email settings controls (Figure 5.5), select **System Settings > Email**.

Figure 5.5. Email settings

The screenshot shows the 'Email Settings' page within the 'System Settings' menu. The page title is 'Email Settings' and it includes instructions: 'Enter a comma-separated list of the email addresses to which reports and alerts should be sent. Leave empty to disable emails.' Below this, the 'Main System Email Settings' section contains several input fields: 'Recipient List (comma separated email addresses, leave blank to disable emails):' with the value 'test@company.com,admin@company.com'; 'Over-ride subject?:' with the value 'WML Email Alert'; 'Sending mail (SMTP) server IP/Network address:' with the value 'smtp.mylan.net'; 'SMTP username (can be blank):' (empty); and 'SMTP password (can be blank):' (empty). An 'Update' button is located at the bottom of the form.

The system may be set up to send reports and alerts to specified email addresses. Enter a comma-separated list of email addresses, or leave the field blank if you do not wish to receive system emails.

You will also need to provide the IP/network address of the SMTP server to be used to send the emails and SMTP username and password, if required.

To save your changes, click on the “Update” button.

Licence key

To access the licence key controls (Figure 5.6), select **System Settings > Licences**.

Figure 5.6. Setting the licence key

Print Platform File Manager System Settings

Network Time Location System Log Email Licences Password

Licence Key

Enter the key to enable your appliance (Mac address = 00:50:f6:2c:86:78)

Enter Key:

Licence valid for:
LICENCE CREDITS: 10
USB DEVICES: 0
ARCHIVE: Print archiving enabled
SHAREINPUT: Network share data input enabled
WMLKYO: WML Prescribe->PDF emulation enabled

Licences in use:
Licence Credits: 2 in use with 8 remaining:
Licence Credit 1: 192.168.0.42
Licence Credit 2: 192.168.0.43
USB Devices: 0 in use with 0 remaining.

Enter the licence key received from the appliance vendor and click on “Update Key”. The licence details will be displayed below.

WML Print Platform Licences have a number of “licence credits” enabled. They also provide a number of flags to enable certain features on the appliance, such as the Prescribe emulation and archiving. A number of USB devices may also be enabled in the licence.

A licence credit enables the use of an IP address on the appliance or the use of a input share/hot folder or archive share (if enabled with the relevant flags). The number of valid licence credits, USB devices and enabled feature flags is shown for the key which is stored on the appliance. The number of licence credits and USB devices which are in use will also be displayed.

System passwords

To access the system password controls (Figure 5.7), select **System Settings > Password**.

Figure 5.7. Setting system and print manager passwords

The screenshot shows a web interface for DataGateway Configuration. At the top, there are navigation tabs: "Print Platform", "File Manager", "System Settings", and "System Settings". Below these are sub-tabs: "Network", "Time", "Location", "System Log", "Email", "Licences", "Password", and "System Settings". The main content area is titled "Password Management" and contains two sections: "Change System Password" and "Change Print Manager Password". Each section includes a note that the password must have a minimum length of 5 characters, followed by input fields for "New Password:" and "Re-type Password:", and an "Update Password" button.

You can use these controls to set passwords for the system and print manager users.

To change either password (you can't change both at once), enter the new password in both boxes and click on "Update Password".

Note

Passwords must have a minimum length of five characters.

The default password is **blankdoc**.

System settings

To access the controls to backup, restore and reset system settings (Figure 5.8), select **System Settings > Backup**.

Figure 5.8. System settings

The screenshot shows the 'System Settings' page with a navigation bar at the top containing 'Print Platform', 'File Manager', 'System Settings', 'Network', 'Time', 'Location', 'System Log', 'Email', 'Licences', 'Password', 'Settings', 'Firmware', and 'Status/Control'. Below the navigation bar, a message states: 'Successfully created the backup file. Now press "Download The Backup" to save a copy onto your system.' The main content area is divided into two columns. The left column, titled 'Create a Backup', includes the instruction 'Choose the selection of settings.' and a list of checkboxes: 'System Files And Settings', 'Network Settings', 'Print Queues' (checked), 'Print Job Data' (checked), 'Overlay And Logic Profiles', and 'Select/Deselect All'. At the bottom of this column are 'Create Backup' and 'Download The Backup' buttons. The right column, titled 'Restore a Backup', includes the instruction 'Upload a previously saved settings file here, then press the "Restore From File" button to apply the settings.' and an 'Upload Backup' button. Below this is the 'Set Factory Defaults' section, which includes the instruction 'Restore factory default settings for passwords and network' and a 'Factory Defaults' button.

Backing up system settings

Before backing up the system settings decide which parts of the configuration need to be stored. The options for back up will include:

- System Files and Settings
- Network Settings
- Print Queues
- Print Job Data
- Overlay and Logic Profiles

There is also a checkbox to select or deselect all options.

It is common to want to copy the print settings from one appliance to another, without making any other changes to the target system. In this case, a backup of "Print Queues" and "Overlay and Logic Profiles" would be a suitable selection.

To copy an entire system setup, but without affecting the network settings of the target appliance, select all options *except* "Network Settings".

To make a full backup to use in the case that the hardware failed and an identical setup were required on new hardware, select all options.

After making the selection, click "Create Backup". A new button "Download Backup" will appear. Click this button and save the file to the local system.

Click on the link provided to backup the system settings. You will be prompted to save a file called "settings.tar". This contains an archive of all the current settings for the device.

Restoring system settings

To restore settings from a previously saved file, click “Upload Backup”, and browse to the file. After the file uploads successfully, a new button “Restore From File” will appear. Click this button to apply the uploaded backup.

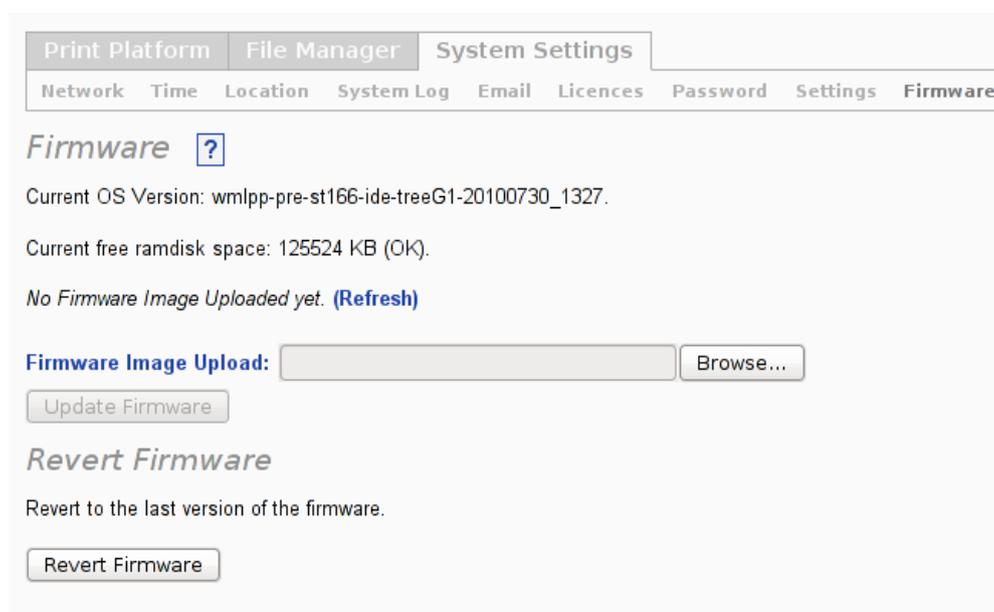
Restoring factory defaults

To restore factory default settings for passwords and network, click on “Factory Defaults”. A confirmation dialog box will appear.

Firmware

To access the system firmware controls (Figure 5.9), select **System Settings > Firmware**.

Figure 5.9. Firmware



To perform a firmware upgrade, you must first upload a new firmware image to the device, either by browsing to the file, by uploading a firmware image into the “update” directory via FTP or by using scp/sftp to copy the firmware image into the “/tmp/uploads/update” directory. Details of the current and uploaded firmware images are shown on the page. Click on “Refresh” if you have uploaded an image and it is not yet displayed.

When the image has been uploaded, the “Update Firmware” button will be displayed. Click on the button to start the update process. The process will reboot the appliance. Avoid interrupting the firmware update, which takes approximately 45 seconds in total. The browser will re-connect to the user interface when the update has completed.

To revert to the previous version of the firmware, click on the “Revert Firmware” button.

Status and Control

The system status and control page provides status information such as system load and memory usage, the ability to shut down or reboot the appliance and a feature to allow WML staff to access the appliance for support.

To access the system status and control page (Figure 5.10), select **System Settings > Status/Control**.

Figure 5.10. Status and Control user interface

The screenshot displays the 'System Status and Control' user interface. At the top, there is a navigation bar with tabs for 'Print Platform', 'File Manager', and 'System Settings'. Below this, a secondary navigation bar includes links for 'Network', 'Time', 'Location', 'System Log', 'Email', 'Licences', 'Password', 'Settings', 'Firmware', and 'Status/Control'. The main content area is titled 'System Status and Control' and is divided into three sections:

- System Status:** A list of system metrics:

System Uptime:	0.944236 hours
System Load Average:	0.31
Persistent Storage:	10 MB
RAMDisk Storage:	122 MB
Total RAM:	247 MB
Active RAM:	25 MB
Cached RAM:	38 MB
Buffered RAM:	19 MB
Inactive RAM:	39 MB
System Log:	View
- System Control:** Two buttons with icons: 'Shut down the system' and 'Reboot the system'.
- Support Access:** A section for enabling support access. It includes the text 'Allow support staff to access this device here.', an unchecked checkbox for 'Enable support access', and the status 'Connection ID: Support access is not running'. Below this, it shows 'Error with support access: None' and two buttons: 'Clear Error' and 'Refresh Status'. A note at the bottom states: 'Support staff will require the connection ID to access your system.'

System Status

This section shows the current system status. Most of the entries in the list are self-explanatory.

“Persistent Storage” is the amount of storage available on the on-board hard disk or flash disk module for storing user files such as PDF forms. When this runs low, the user will be unable to add new PDF forms to the appliance.

The “RAMDisk Storage” is used for all temporary files and is half the size of the total installed RAM. If this runs low, then print jobs may stop processing through the DataGateway. During the implementation phase, it is important to ensure that the maximum print load will not cause the RAMDisk Storage to run out.

The “View” button for the “System Log” will display a pop-up window showing the last few lines of the internal system log (which resides on the RAMDisk and is not persistent across reboots).

The “View” button for “Network Mounts” displays a window showing mounted network shares. This is used to verify that any archive and hot folder shares which have been con-

figured have been successfully mounted by the appliance and have sufficient storage available.

System Control

This page provides shut down and reboot buttons.

Shut down

The shut down button will display a confirmation dialog box. After confirmation has been given, the system will go through a safe shut down procedure. The web interface will display a success message when it is no longer able to connect to the appliance's embedded web server.

Reboot

If the user confirms a reboot, the appliance will safely shut down, then reboot. The web interface will re-connect to the appliance login screen as soon as the reboot is complete.

Support Access

The DataGateway provides a remote access mechanism to allow support staff to access the appliance and investigate problems.

To use this feature, the appliance must have internet access. If a firewall is in use, it must be configured to allow the appliance to initiate a direct connection on TCP port 80 (HTTP) to the support access server (rssh.wmltd.co.uk). This is equivalent to allowing the appliance to browse the web.

To turn support access on, check the box "Enable support access". The appliance will attempt to connect to the support access server. When a successful connection is made, a connection ID will be displayed. Support staff will request this connection ID to enable them to connect back to the appliance.

Chapter 6. Queue Management

Queue management is carried out in the **Print Platform** section of the DataGateway user interface. This section contains the queue management interface, comprising the **Queues** page, which lists and monitor all queues on the appliance, and the **Queue Config** section, where detailed configuration of queues can be carried out.

Other pages in this section are the **Jobs** page, on which recent jobs are listed and managed, and the **Platform Settings** page, on which platform-wide settings may be configured.

6.1. Queue management interface

Simple queue administration tasks, such as adding a queue or changing input and output details, may be carried out on the **Print Platform > Queues** page. Full configuration is carried out in **Print Platform > Queue Config**, described in Section 6.2.

The first point of entry to the queue management interface is the print queues list (Figure 6.1), found on the **Print Platform > Queues** page.

Figure 6.1. The print queues list



This list displays and monitors all queues on the DataGateway. The summary view shown in Figure 6.1 displays a status line for each queue, comprising the queue name, activity monitor and queue status information, as well as a number of print user actions for the queue. The queue list is updated periodically so that the information shown always reflects the current state of the queues.

System status information is displayed above the queue list. The system load average and the total and inactive memory are displayed. The information is updated periodically to show the current values.

Figure 6.2 details the components of the queue status line.

Figure 6.2. The queue status line

Print Queues

This system has 25 licence credits with 12 in use.

System status

System Load Average: 1.72

Total/Inactive RAM: 501 MB/58 MB

Queue Name	Info	Activity	Status	Last Job	Print	Jobs	Status
tq	1	2	3	idle	stopped at 2012/02/08 11:51:44	4	
tqfinal				idle	completed at 2012/02/28 10:23:23		
tqpcl				idle			
tqps				idle			
tqsimple				idle	completed at 2012/02/17 17:02:07		

- ❶ Queue info button. Click to display a summary of the queue configuration.
- ❷ Activity monitor. The bars light up when jobs are processing on the queue.
- ❸ Queue status. The current status of the queue and information about the last job processed on the queue.
- ❹ Print user actions: print file, view jobs and print status page.

Print user actions

The queue status line enables a number of print user actions, which are described below.

Print file

Click on the print icon (🖨️) to send a file to print on the queue. You will be prompted to browse to a file. Once a file has been selected, it will be sent to print.

View jobs

The view jobs icon (📄) provides a direct link to the **Print Platform > Jobs** page (see Section 6.4), set to display all jobs on the selected queue.

Print status page

Click on the print status icon (📄) to send to print a queue status report page (Figure 6.3) providing queue details and status information and current system information.

Figure 6.3. Queue status report page

WML **Queue Status Report**
2012-02-28_15:48:16

Queue Information

Name: lq
Info: Test kyo queue
Location:
Make/Model: Kyocera Mita FS-9500 Foomatic/Is9500
Device URI: lpd://127.0.0.1/lq/hal
Members (if applicable):

Queue Status

Status: idle [shared]
Last Job: stopped at 2012/02/08 11:51:44
Enabled?: yes
Accepting?: yes

System Information

Firmware: wmlpp-dg-kvm-sata-tree1-20120228_0954
Hostname: wmlpp **Total Memory:** 501 MB
IP address: 192.168.0.122 **Active Memory:** 95 MB
Mac address: 52:54:00:67:54:7e **Buffered Memory:** 39 MB
Load Average: 1.57 **Inactive Memory:** 61 MB

Printed at 3:48 PM 28/02/2012

Innovative Technology

Print admin controls

The print queues list provides admin controls which allow simple administrative tasks to be carried out on existing queues and which facilitate the addition of new queues.

The print admin controls for an existing queue are revealed by clicking on the status line for the queue. The controls will appear in a new line below the queue status line, and labels for the controls will appear in the queue list header (Figure 6.4). Click on the status line again to hide the controls.

Tip

To show or hide the print admin controls for all queues, click on the queue list header.

Figure 6.4. Print admin controls

Print Queues

This system has 25 licence credits with 12 in use.

System status
System Load Average: 1.72
Total/Inactive RAM: 501 MB/58 MB

Queue Name	Input Port	Info	Output Protocol	Destination Network Name or IP Address	Output Port/Queue Name	Action	Select
tq <input type="text" value="tq"/>	<input type="text" value="9100"/>	<input type="text" value="Test queue"/>	<input type="text" value="lpd"/>	<input type="text" value="127.0.0.1"/>	<input type="text" value="tqfinal"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> Enable <input checked="" type="checkbox"/> Accept Jobs <input checked="" type="checkbox"/> Shared							
tqfinal <small>Status: idle Last Job: stopped at 2012/02/08 11:51:44</small>							
tqpcl <small>Status: idle</small>							
tqps <small>Status: idle</small>							
tqsimple <small>Status: idle Last Job: completed at 2012/02/17 17:02:07</small>							
Add a new queue							
<input type="text"/>	<input type="text" value="9109"/>	<input type="text"/>	<input type="text" value="lpd"/>	<input type="text"/>	<input type="text" value="passthru"/>	<input type="checkbox"/>	<input type="checkbox"/>

Note

Please note that **Info** field is hidden at narrow browser widths. For access to this field, your browser window should be 1000px or more wide.

Editing a queue

Changes made to the queue name take effect as soon as focus leaves the queue name field. Modifications made to other fields should be saved by clicking the update button (↻).

Note

The queue name may contain upper and lower case letters, numbers, the dash character and the underscore character. Queue names may not contain spaces.

Click on the configure button (⚙) to gain access to the full queue configuration interface for the queue.

A queue may be duplicated by clicking on the duplicate button (📄). This will create a copy of the queue, including full configuration details. You will be given the opportunity to edit the automatically-generated name for the new queue before the new queue is created.

An existing queue may be removed by clicking on the delete button (✖). You will be prompted to confirm the deletion. You can also delete multiple queues by checking the “Select” checkbox for each queue, then clicking on the delete selected button (✖).

Adding a print queue

To add a queue, fill in the queue name, input port, output protocol, destination network name or address and output port/queue name in the “Add a new queue” area, then click on the add button (+). The queue will be added to the system.

Note

The queue name may contain upper and lower case letters, numbers, the dash character and the underscore character. Queue names may not contain spaces.

The input port defaults to the next available port number, starting at 9100, a value which is commonly used in printing applications. It is possible for two queues to respond on the same raw port. In this case the first queue (alphabetically) will respond to incoming print data on the shared port, **unless you have enabled per-IP address filtering in the input tab of the queue configuration**. To find out more about how queues can respond based on sender and destination IP addresses, see the section called “LPD/raw input settings”.

Classes

The DataGateway can be configured with *print classes*. A print class is a group of print queues. Print data sent to a print class will be forwarded to members of the class in rotation.

The classes interface is similar to that for print queues (Figure 6.5).

Figure 6.5. The classes list



To add a class, enter a name for the class, and select at least one member for the class. Multiple queues may be selected by holding down the control key. The input port defaults to the next available port number, starting at 10100. Click on the add button (+) to add the class to the system.

Note

At least one queue must be configured before it is possible to add a class.

Click on the configure button (⚙️) to gain access to the full configuration interface for the class. Classes have **Input** and **Output** options, but no data processing options; all data processing is configured on a per-queue basis.

6.2. Queue configuration

While some queue parameters can be modified in the print queue list, full configuration options for each queue are reached via the configure button in the print queues list, as shown in Figure 6.6, or by going to the **Print Platform > Queue Config** page and switching to the required queue using the **Switch to queue** menu.

Figure 6.6. The configure button in the print queues list

The screenshot displays the 'Print Queues' management page. At the top, it shows system status: 'System Load Average: 1.72' and 'Total/Inactive RAM: 501 MB/58 MB'. Below this, a table lists existing queues:

Queue Name	Input Port	Info	Output Protocol	Destination Network Name or IP Address	Output Port/Queue Name	Action	Select
tq	9100	Test queue	lpd	127.0.0.1	tqfinal	[Configure] [Refresh] [Delete]	<input type="checkbox"/>
tqfinal						[Configure] [Refresh] [Delete]	<input type="checkbox"/>
tqpcl						[Configure] [Refresh] [Delete]	<input type="checkbox"/>
tqps						[Configure] [Refresh] [Delete]	<input type="checkbox"/>
tqsimple						[Configure] [Refresh] [Delete]	<input type="checkbox"/>

Below the table is an 'Add a new queue' section with input fields for Queue Name, Input Port (9109), Output Protocol (lpd), Destination Network Name or IP Address, and Output Port/Queue Name (passthru).

The following sections describe the options available for configuration within the **Print Platform > Queue Config** section.

Data input

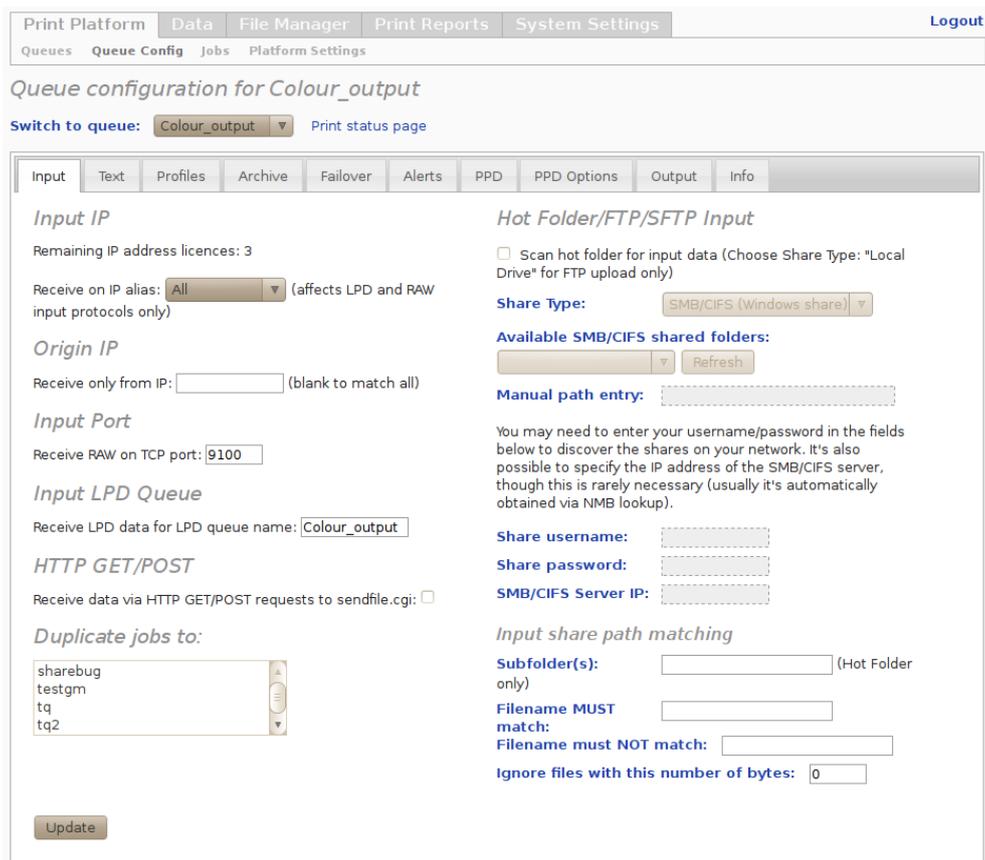
Print data can be sent to the DataGateway using a variety of methods. These include the well-known Berkeley *LPD protocol*, defined in Internet RFC 1179, the *IPP protocol*, which is defined in RFC 3510 and the *raw print protocol*, by which data is sent unmodified, with no additional command metadata, to a particular TCP port.

In addition to these input protocols, the DataGateway can receive data from an *HTTP POST request*, where the print data is encoded in a POST form and it can also collect print files from a network-shared hot folder. This is called the *input share* or *hot folder* method.

The DataGateway also has a unique feature which allows it to behave as if it were any number of printers on a network, each using the same TCP port for raw printing. The DataGateway can be configured with a number of IP address aliases. Any queue can then be instructed to “listen” only on a particular IP address, ignoring input to other IP addresses.

Select the **Input** tab (Figure 6.7) to configure the input data options for the queue.

Figure 6.7. Queue configuration: Input tab



The left-hand column contains the input settings for LPD/raw input data. The right-hand column contains the input share settings.

Caution

Before changing any settings, shut down any print job scheduler which may be sending data to the DataGateway. If print jobs are spooling onto the DataGateway when the queue settings are updated, data loss may occur.

LPD/raw input settings

The input IP on which to listen for raw print data can be selected from the “Receive on IP alias” drop-down control. This shows the available IP addresses on which the device can

listen, including all the aliases. To add aliases, use the Network settings page in **System Settings > Network**.

It is possible to filter jobs using the *origin IP*. Enter an IP address in the **Receive only from IP** field and the queue will only accept jobs originating from that single IP address.

Note

It is only possible to specify a single IP address or “all” for the input IP and the origin IP. To process data from two or more originating servers, create additional print queues with identical settings and set input/origin IP settings for the two or more addresses.

The **Input Port** for raw TCP data input can be set to a value in the range 1024 to 32767, inclusive.

Note

The input port can also be set in the Print Queues list.

Note

If the CDL module has been licenced (DataGateway only) some additional fields will be visible here, allowing you to specify that the incoming data is expected to be CDL-formatted data.

The **Input LPD Queue** field can be set to a value with less than 256 characters and no spaces. This alias can be used with the LPD input protocol in place of the actual name of the queue. The LPD queue alias defaults to the name of the queue.

HTTP GET/POST input

To enable the queue to receive data via an HTTP GET or POST request to **sendfile.cgi**, check the box in **HTTP GET/POST**.

Note

The HTTP GET/POST input feature uses a licence credit for each queue on which it is enabled.

Making an HTTP POST request

Note

While it is possible to use an HTTP GET request to send data to the DataGateway, it is more natural to use HTTP POST, which is the recommended request to use.

To print some data to the DataGateway, make an HTTP POST request with the following parameters in POST form (use the application/x-www-form-urlencoded POST request type):

- queue: The queue to which you wish to send the data (required).
- copies: The number of copies of the data which should be output (defaults to 1 if omitted).
- title: The title for the job (optional).
- user: The user who sent the data (optional).
- data: The data for the job (required).

All parameters should be url-encoded. This will be necessary for most of the fields; certainly for the data.

Make the request to **http://your.DataGateway.net/sendfile.cgi**. An XML-formatted response will be returned.

A useful utility for testing is *curl*, available, at the time of writing, from <http://curl.haxx.se/>.

The following call to the curl command will spool one copy of the file "file.name" giving it the title "test", and printing it to the queue "testgm" on the DataGateway "wml.localnet" on behalf of the user "Bob":

```
curl -X POST http://wml.localnet/sendfile.cgi -d queue=testgm -d copies=1 \
-d title=test -d user=Bob --data-urlencode data@file.name
```

Note

If copies is set to a number N, greater than 1, then the data will be spooled as *input* N times, rather than having the output of any processing on the queue emitted N times. Any processing which the DataGateway would carry out on the data will be repeated N times.

The response to an HTTP POST request

The response returned by call to **sendfile.cgi** is XML data which indicates the success or failure of the request.

If the request was successful, the XML will contain an element called **success** meaning that the data was accepted onto the queue. It doesn't mean that the data was successfully *output*, only that the data was received and accepted onto the queue.

If there was an error which prevented the data from being accepted, then a field called **error** will be found in the XML response. This error message will describe what went wrong.

In addition to the success and error elements in the XML, the parameters of the request will be returned back to the caller, which can be helpful when testing the process.

An example of a successful response is shown below. The **success** element contains the message "Successfully spooled data to" followed by the target queue name (testgm, in the example).

```
<?xml version="1.0" ?>
```

```
<!DOCTYPE WmlSendFileResponse>
<WmlSendFileResponse>
  <copies>1</copies>
  <dataSizeBytes>674</dataSizeBytes>
  <jobId>34933</jobId>
  <queue>testgm</queue>
  <success>Successfully spooled data to testgm</success>
  <title>test</title>
  <user>seb</user>
</WmlSendFileResponse>
```

In the next response, the call failed because the user did not pass a data file:

```
<?xml version="1.0" ?>
<!DOCTYPE WmlSendFileResponse>
<WmlSendFileResponse>
  <dataSizeBytes>0</dataSizeBytes>
  <error>Please pass document data in the parameter "data".</error>
</WmlSendFileResponse>
```

The full list of elements in the response XML

- **copies:** The number of times to spool the data to the target queue.
- **dataSizeBytes:** The number of bytes of data presented in the POST (or GET). Always returned, whether request is a success or not.
- **error:** An error message. If present, the data was not accepted onto the queue.
- **jobID:** The internal CUPS job ID for the successfully spooled data. Only present if the job is successfully spooled. If copies was set to >1, then this will contain a comma-separated list all IDs.
- **queue:** The target queue specified in the request.
- **success:** A success message. Only present if data is successfully accepted.
- **title:** The job title specified in the request.
- **user:** The user specified in the request.

The most common HTTP POST request error messages

- Please pass document data in the parameter "data". *Check your HTTP request.*
- The target queue does not accept HTTP GET/POST requests to spool data. *Make sure you have enough licence credits and ensure that HTTP GET/POST is enabled in **Print Platform > Queue Config > Input** .*
- Please set the target queue in the parameter "queue". *Check your HTTP request.*
- The queue is not accepting jobs. *Check that the **Accept** box is checked for your target queue in **Print Platform > Queues**.*
- The print transport (CUPS) did not accept the job. *Check you haven't used up all the ramdisk or persistent storage on the system (**System Settings > Status/Control**). Try re-starting CUPS at **Print Platform > Platform Settings > Cups Control**.*

Duplicating input jobs

The **Duplicate jobs to:** box is a multi-select menu. Any queues highlighted in this box will receive a copy of every input data file received on the current queue. Hold the Ctrl key to select multiple queues in this box.

Share-based input

To print files from a network share, the DataGateway first “mounts” the network share, then scans it on a regular basis (every second) for new files to print. When it finds a file, it monitors its size, then when the size has been static and non-zero for 3 seconds (the “settle-time”), it spools it for print.

Empty files (with zero size) are not printed. The DataGateway logs the existence of empty files in the input share with a message like this:

```
Oct 16 10:08:35 wmldevice ctrl1[731]: \  
    Ignoring zero sized file: '/tmp/hotfolder/myqueue/file.pdf'
```

This log message is repeated every 60 seconds.

Caution

The input share settle-time is only 3 seconds.

Files *must* be fully copied onto the input share within this time-scale.

Care must be taken with the share-based input method if your application creates files on the input share using file access programming techniques.

If you use functions such as `fopen()` in C or file-streams in C++, you must ensure that your application creates the file in less than the settle-time of 3 seconds.

Suppose your application creates a file in the input share, sends *some*, but not *all* of the print data into the file, then pauses for more than the settle-time before completing. The DataGateway will see the the file size is non-zero and has remained static for 3 seconds, so it will pick up the incomplete job, and attempt to print it.

To avoid this problem, you can use two strategies: Either have your application generate the file in a separate folder and then copy the completed file into the DataGateway input share *or* have your application create each file with a “.spl” suffix (meaning SPooling). The DataGateway will ignore files with a “.spl” suffix (which must be in lower case). When your application has completed generating the file, have it re-named without the .spl suffix. The job will then be printed.

Once printed, the DataGateway erases the file from the network share. For this reason, the network share used for printing should be configured for read/write access and if a copy of the print data is required it should either be archived using the DataGateway's archiving features, or archived before it is dropped in the input share.

To enable share-based input, check the box labelled **Scan share directory for inputs**. It's possible to use either Unix/Linux Network File Server shares or Windows/Samba SMB/CIFS network shares from which to collect print data.

Input Share (Hot Folder)

Scan hot folder for input data

Share Type:

Available SMB/CIFS shared folders:

Manual path entry:

You may need to enter your username/password in the fields below to discover the shares on your network.

Share username:

Share password:

Selecting a share

NFS (Linux/Unix shares)

To connect to a Linux/Unix NFS share, choose **NFS (Linux/Unix)** from the **Share Type** drop-down. Enter the path to the NFS share in the following format:

[Network/IP Address]:[folder path]

For example:

192.168.100.100:/input_data

The NFS server's exports file should be configured to allow both read and write access to the directory path (/input_data in the example). An example of a suitable line in /etc/exports on the NFS server is:

```
/input_data 192.168.0.0/255.255.255.0(rw,no_root_squash, sync, subtree_check)
```

If the NFS share is to be used as an *archive* Share, then correct permissions must be set on the share directory on the NFS server. The following would be required for an archive share specified as:

192.168.100.100:/archiveData

On the server it would be necessary to change the user on this directory to “root” or 0 and the group of the directory to 7, which is the GID for the “lp” user on some systems:

```
root@nfsServer# chown 0:7 /archiveData
```

Finally, ensure that both user and group are able to read and write to the directory:

```
root@nfsServer# chmod ug+rwX /archiveData
```

SMB/CIFS (Windows shares)

For Windows SMB/CIFS network shares, select **SMB/CIFS (Windows share)** and click on the **Refresh** button to scan the network for available shares. This will populate the **Available SMB/CIFS shared folders** drop-down.

Important

Most recent Microsoft Windows file servers will not broadcast their available shares to the network without authentication.

If necessary, enter your network username and password in the fields **Network user name** and **Network password** and then press **Refresh**. The username and password will be used to obtain the list of available shares.

It is also possible to enter the share path manually in the **Manual path entry** text field, by choosing the option **Enter path in text box** from the **Available SMB/CIFS shared folders** drop-down.

Enter the SMB/CIFS path in the format:

`\\MACHINE\shareName`

The **Network user name** and the **Network password** fields must be filled in to attach shares which require authentication.

Unsupported share types

The DataGateway print platform uses a subset of the tools from Samba (www.samba.org) to scan for, and attach, SMB/CIFS network shares. Some Microsoft Windows share configurations are not supported and cannot be used for input share access on the DataGateway. In particular, Active Directory shares cannot be used. There may also be problems with certain WINS configurations.

Overlay templates and logic profiles

The profiles tab allows you to “bind” an overlay template or a logic profile to the queue.

An overlay template is a series of instructions to be applied to the print data passing through a queue, such that the resultant print job may be overlaid with a set of fixed or variable data texts, or the contents of one or more PDF files.

A logic profile is a set of rules to apply to the data stream. Logic profiles scan the text in the data stream and apply “logic actions” where conditions are matched.

To change the overlay template for the queue, select the **Profiles** tab (Figure 6.8).

Figure 6.8. Queue configuration: Profiles tab

The screenshot shows the 'Profiles' tab in the Queue Management interface. At the top, there are navigation tabs: 'Print Platform', 'Data', 'File Manager', 'Print Reports', and 'System Settings'. Below these are sub-tabs: 'Queues', 'Queue Config', 'Jobs', and 'Platform Settings'. The main heading is 'Queue configuration for Default'. A 'Switch to queue:' dropdown is set to 'Default', with a 'Print status page' link. The 'Profiles' sub-tab is active, showing two sections: 'Overlay Profile' and 'Logic Profile'. The 'Overlay Profile' section includes instructions on choosing an overlay profile and selecting 'Triggered' for dynamic switching. It features a 'Use overlay profile:' dropdown set to 'Advert Demo' and an 'Update' button. The 'Logic Profile' section includes instructions on choosing a logic profile and a 'Use logic profile:' dropdown set to 'None'.

You may choose to apply an existing overlay template to the queue by selecting the template name from the drop-down control. Alternatively, if you wish to switch overlay set dynamically using the input data for a print job, you can select **Triggered**. The print input data should then include the name of the overlay template to be used.

Logic profiles are bound to the queue in a similar way using the corresponding drop down menu of available logic profiles.

Note

If both an overlay template *and* are bound to a queue, then any overlay template which is triggered by the logic profile will be applied to the job. If no template is triggered by the logic profile, the template selected here will be applied.

Configuration of overlay templates and details about the structure of input data are described in Chapter 7. Logic profiles are described in Chapter 8.

Archiving

A queue may be configured to archive print jobs to network attached storage. The network share may be either a Unix/Linux NFS share or a Windows/Samba SMB/CIFS network share.

To configure the archive settings for the queue, select the **Archive** tab (Figure 6.9).

Figure 6.9. Queue configuration: Archive tab

Queue configuration for SecureDelivery

Switch to queue: [Print status page](#)

Input Text Profiles **Archive** Failover Alerts PPD PPD Options Kyo Settings Output Info

Archiving

Archive jobs to network attached storage

Share Type:

Available SMB/CIFS shared folders:

Manual path entry:

You may need to enter your username/password in the fields below to discover the shares on your network.

Share username:

Share password:

Electronic Delivery

Deliver electronically:

Archive expiry: days

Archive File Name

The archive file name will be formatted as follows. Use the controls below to customize the format.

Fields

Queue
Sub-folder
Date: Feb2112
Time: 115115
Job title

Add a field

Field separator:

Suffix:

To turn on archiving, check the **Archive jobs to network attached storage** box. This will activate the **Share Type** drop-down and other form controls, allowing you to select the storage to be used for the archive.

Selecting a share

NFS (Linux/Unix shares)

To connect to a Linux/Unix NFS share, choose **NFS (Linux/Unix)** from the **Share Type** drop-down. Enter the path to the NFS share in the following format:

[Network/IP Address]:[folder path]

For example:

192.168.100.100:/input_data

The NFS server's exports file should be configured to allow both read and write access to the directory path (/input_data in the example). An example of a suitable line in /etc/exports on the NFS server is:

```
/input_data 192.168.0.0/255.255.255.0(rw,no_root_squash, sync, subtree_check)
```

If the NFS share is to be used as an *archive* Share, then correct permissions must be set on the share directory on the NFS server. The following would be required for an archive share specified as:

192.168.100.100:/archiveData

On the server it would be necessary to change the user on this directory to “root” or 0 and the group of the directory to 7, which is the GID for the “lp” user on some systems:

```
root@nfsServer# chown 0:7 /archiveData
```

Finally, ensure that both user and group are able to read and write to the directory:

```
root@nfsServer# chmod ug+rwX /archiveData
```

SMB/CIFS (Windows shares)

For Windows SMB/CIFS network shares, select **SMB/CIFS (Windows share)** and click on the **Refresh** button to scan the network for available shares. This will populate the **Available SMB/CIFS shared folders** drop-down.

Important

Most recent Microsoft Windows file servers will not broadcast their available shares to the network without authentication.

If necessary, enter your network username and password in the fields **Network user name** and **Network password** and then press **Refresh**. The username and password will be used to obtain the list of available shares.

It is also possible to enter the share path manually in the **Manual path entry** text field, by choosing the option **Enter path in text box** from the **Available SMB/CIFS shared folders** drop-down.

Enter the SMB/CIFS path in the format:

```
\\MACHINE\shareName
```

The **Network user name** and the **Network password** fields must be filled in to attach shares which require authentication.

Archive file name format

The format for the file names of archived print jobs can be controlled in the **Archive File Name** section of the **Archive** tab. An example file name is shown in the box, and the controls below can be used to customize the file name format.

The **Fields** box displays the list of fields used to make up the current archive file name. Add fields by choosing them from the dropdown menu and clicking on the add button (+). You can re-order or delete fields by highlighting them and clicking on the up and down arrows or delete button. You can also choose the field separator character and the file suffix to be used. The example file name will be updated to reflect your changes.

To save your changes, click on the **Update** button.

Archive expiry

You can choose to have archive files "expire" after a given number of days. This means that they will be removed from the system after the given time period has elapsed. If the **Archive expiry** value is set to 0, then archive files will not expire.

Electronic delivery

If the Electronic Document Delivery module is enabled, then a checkbox to set the queue's output to be delivered electronically will be present.

Note

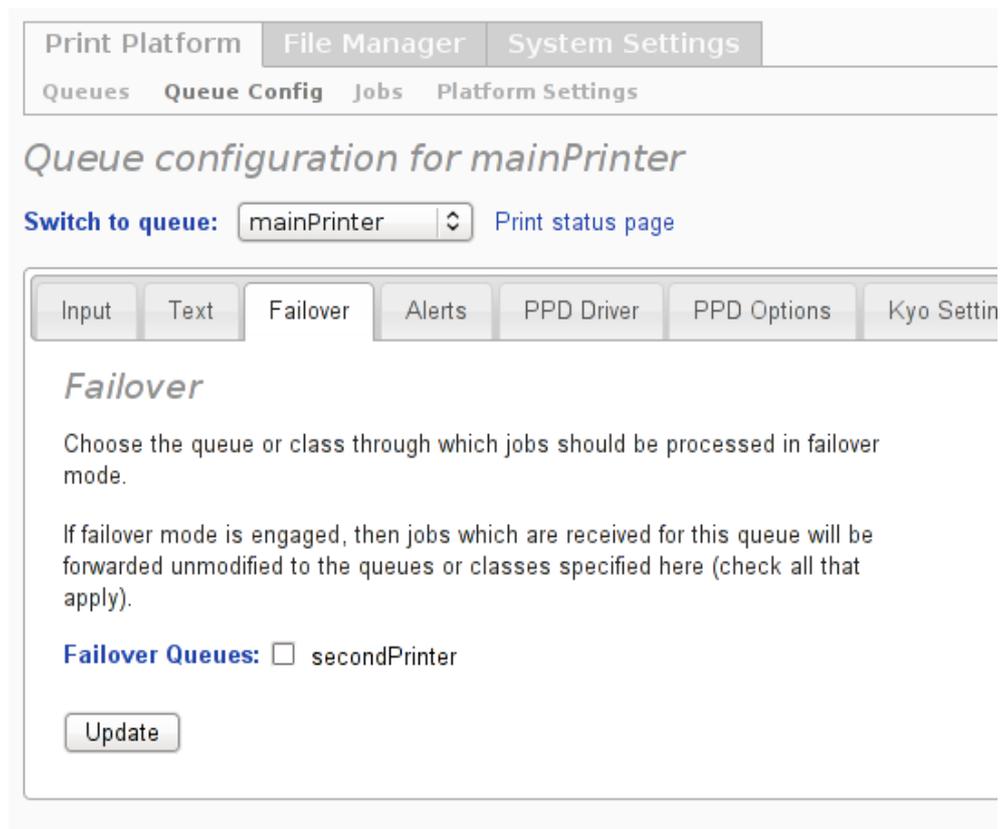
The archive feature of DataGateway is used to make a store of the documents for electronic delivery. For this reason, archiving must be enabled on the queue for electronic document delivery to work.

Failover

The DataGateway can be placed into *failover mode* in **Platform Settings > Failover**. In failover mode, incoming print jobs will be passed to the “failover queue(s)”.

To set the failover queues for the current queue, select the **Failover** tab (Figure 6.10).

Figure 6.10. Queue configuration: Failover tab



You may select the queues through which jobs should be processed in failover mode.

If failover mode is engaged, then jobs which are received for the current queue will be forwarded unmodified to the queues specified.

If **Pool** is selected, then jobs will be failed over to each of the selected queues in turn.

To save your changes, click on the **Update** button.

Alerts (Queue Monitoring)

The DataGateway system may be configured to send alerts if the queue does not appear to be processing print jobs at the expected rate. The alerts are sent as emails, using the main system email settings configured in **System Settings > Email**, or per-queue settings, if provided.

Alerts can be sent when too few documents are processed, on a regular basis with a status message about how many jobs have been processed, or for every individual document processed on the queue.

To configure the queue alert settings, select the **Alerts** tab (Figure 6.11).

Figure 6.11. Queue configuration: Alerts tab

The screenshot shows the 'Alerts' tab in the 'Queue configuration for mainPrinter' interface. At the top, there are navigation tabs: 'Print Platform', 'File Manager', 'System Settings', 'Queues', 'Queue Config', 'Jobs', and 'Platform Settings'. Below these, the title 'Queue configuration for mainPrinter' is displayed, followed by a 'Switch to queue:' dropdown menu set to 'mainPrinter' and a 'Print status page' link. The 'Alerts' tab is selected, with other tabs like 'Input', 'Text', 'Failover', 'PPD Driver', 'PPD Options', and 'Kyo Se' visible. The main content area is titled 'Alerts' and contains the following text: 'Configure the "queue alert" timings here.' 'Set the working hours during which the queue should be monitored. This time is used daily, but Saturday and/or Sunday can be excluded using the checkboxes.' 'Finally, enter a threshold time (in minutes) and a threshold number of print jobs. If, after the threshold time has expired, fewer than the threshold number of jobs have been printed on the queue, an alert email will be sent out to the recipients listed in [System->Emails](#)'. Below this text are two time selection controls: 'Start Monitoring at: 00:00' and 'stop Monitoring at: 00:00' (with a note '(Set both to 00:00 to monitor continuously)'). There are two checkboxes: 'Exclude Saturday Monitoring: ' and 'Exclude Sunday Monitoring: '. A note says 'Set either time or jobs to 0 to disable monitoring:'. Below that are two input fields: 'Time Threshold (minutes): 0' and 'Minimum Jobs Threshold: 0'. At the bottom, there is a checkbox 'Send email even when enough jobs are processed: ' and an 'Update' button.

Firstly, set the working hours during which the queue should be monitored using the drop-down controls. Choose a start time and an stop time for the queue monitoring period. This monitoring period will be applied daily, though Saturday and Sunday can optionally be excluded using the checkboxes for five and six day production weeks.

To monitor the queue continuously, set both **start monitoring** and **stop monitoring** times to **00:00**.

Next, enter a threshold time (in minutes) and a threshold number of print jobs. If, after the threshold time has expired, fewer than the threshold number of jobs have been printed on the queue, an alert email will be sent out to the recipients listed in **System Settings > Email** (see “Email settings” in Chapter 5: *DataGateway Configuration*).

Set either **Time Threshold** or **Minimum Jobs Threshold** to **0** to disable monitoring.

If a status message is required even when the minimum threshold has been exceeded, check the **Send monitoring email even when...** checkbox.

To configure the system to send an alert email for every single job processed during the monitoring working hours, check the box **Send monitoring email for EVERY job...**

To save the queue alert settings, click on the **Update** button.

PPD driver

To select a PPD driver for the queue, select the **PPD Driver** tab (Figure 6.12).

Figure 6.12. Queue configuration: PPD Driver tab

The screenshot shows a web-based configuration interface for a printer queue. At the top, there are navigation tabs: 'Print Platform', 'File Manager', and 'System Settings'. Below these are sub-tabs: 'Queues', 'Queue Config', 'Jobs', and 'Platform Settings'. The main heading is 'Queue configuration for mainPrinter'. A dropdown menu shows 'mainPrinter' selected, with a 'Print status page' link next to it. Below this is a row of tabs: 'Input', 'Text', 'Failover', 'Alerts', 'PPD Driver' (which is active), 'PPD Options', 'Kyo Settings', and 'Output'. The 'PPD Driver' section shows the current PPD file as 'HP 910, hpcups 3.9.12' with a document icon. Below this is the heading 'Add/Modify Driver PPD' and a brief instruction: 'Either choose a new PPD driver from the list or upload a new PPD file. PPD files describe the options available on a particular printer.' There are two dropdown menus: 'Choose printer make:' set to 'HP' and 'Choose printer model:' set to 'HP 910, hpcups 3'. An 'Update' button is below these. Below the 'Update' button is the word 'or'. At the bottom, there is a label 'Add a PPD file or package:' followed by a text input field and a 'Browse...' button.

A PPD file is a *printer description*. It describes the options available for a particular printer and what data format the printer requires (PostScript, PCL, raster data etc). It's not actually a printer driver, but it does specify which driver to use, and what the parameters of the print driver should be.

The current PPD file is displayed at the top of the page. You can view the content of the file by clicking on the file name.

A selection of PPD files are shipped on the DataGateway. These are available for selection in the “Choose printer make” and “Choose printer model” menus. First choose the make of your printer, then a selection of models will be listed in the “Choose printer model” menu.

It is also possible to upload an individual PPD file or a package of PPD files via the “Add a PPD File or package” file upload dialog.

If an individual PPD file is uploaded, it will be stored (and added to the make/model menus) and set as the current PPD.

A package of PPD files is added to the make/model menus. PPD packages can be rpm packages, deb packages, zip archives or tar archives. These are automatically uncompressed and unpacked.

All uploaded PPD files are listed in **File Manager > PPD Files**

When a PPD file is set for a queue, the driver options for the printer will be made available (see “Driver options”).

Internal data PPD files

The DataGateway contains PPD files which force print data output to be PDF, generic PCL or “raw”, where not modification is made to input print data. These PPDs all have the make “WilliamMatthew”.

Driver options

To modify printer driver options for the queue, select the **PPD Options** tab.

Printer driver options are displayed here. The options will match the current PPD file (see “PPD driver”).

Figure 6.13. Queue configuration: Driver Options tab

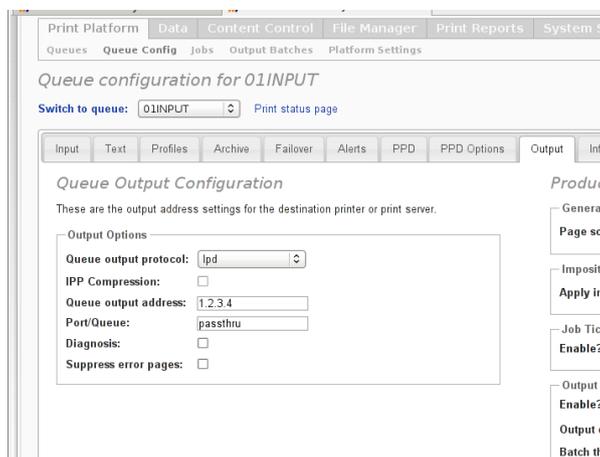
The screenshot shows the 'Queue configuration for mainPrinter' interface. At the top, there are navigation tabs: 'Print Platform', 'File Manager', and 'System Settings'. Below these are sub-tabs: 'Queues', 'Queue Config', 'Jobs', and 'Platform Settings'. The main title is 'Queue configuration for mainPrinter'. Below the title, there is a 'Switch to queue:' dropdown menu set to 'mainPrinter' and a 'Print status page' link. The main configuration area has several sub-tabs: 'Input', 'Text', 'Failover', 'Alerts', 'PPD Driver', 'PPD Options' (which is selected), and 'Kyo Settings'. Under the 'PPD Options' tab, there are two sections: 'General' and 'Installable Options'. The 'General' section includes: 'Media Size' (Letter 8.5x11 in), 'Media Size' (Letter 8.5x11 in), 'Double-Sided Printing' (Off), 'Media Source' (Upper Tray), 'Output Mode' (Color), 'Media Type' (Plain Paper), and 'Print Quality' (Best). The 'Installable Options' section includes: 'Duplexer Installed' (Not Installed). At the bottom left of the configuration area is an 'Update' button.

Use the drop-down controls to modify the driver options as required. To save your changes, click on the “Update” button at the bottom of the list of options.

Output

The output options for a queue are configured in the **Output** tab (Figure 6.14).

Figure 6.14. Queue configuration: Output tab



Basic output options

For lpd, raw, ipp, and ipps protocols, select the required output protocol from the menu and enter a suitable destination address (Queue output address) and port or queue.

IPP options

If either **ipp** (Internet Printing Protocol) or **ipps** (Secure IPP) is chosen for **Queue output protocol**, it is possible to specify whether IPP compression should be applied to outgoing data. Use of this feature requires that the receiving IPP device is able to decompress the data. If the receiving IPP device is not able to apply decompression, then the data will be sent without compression.

Generally IPP print jobs are sent to port 631. If a different port is required, do not enter it in **Port/Queue** - this is where the IPP destination queue name must be entered. Instead, enter the port with the address in the following format:

```
192.168.0.100:632
```

Queue diagnostics

For diagnostic purposes, the print job can be saved before and after it is processed by the DataGateway software. The input data will appear in the file list as prefilter_queue.bin and the output as postfilter_queue.bin.

Enable diagnostics by checking the **Diagnosis** checkbox.

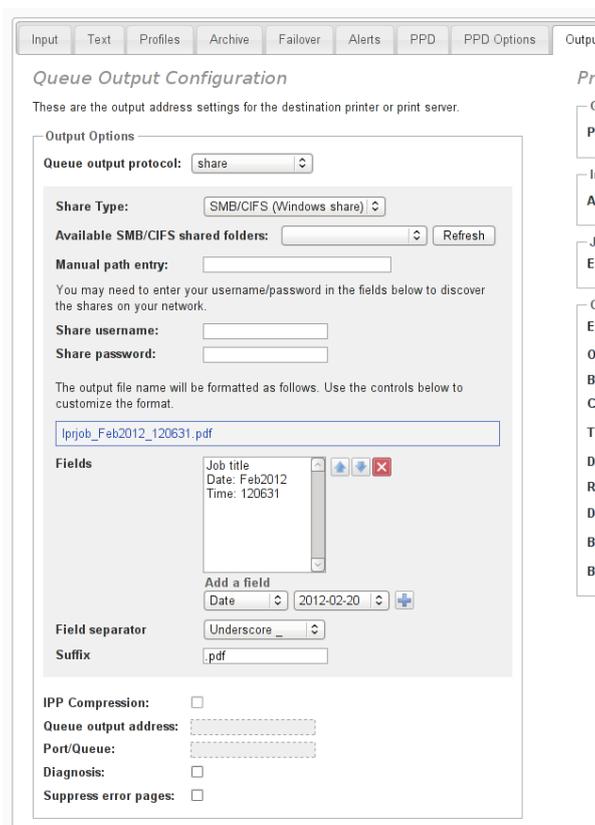
This feature has the potential to fill the storage of the device and lead to failed print jobs, so care must be taken with its use. Because of this, the feature will always be switched off following a reboot of the DataGateway.

Suppress error pages

Many WML features will generate a queue error page to explain the reason why it can't complete the print job. In certain cases it may be useful to suppress these pages. To do so, check the box **Suppress error pages**.

Output to network share

Figure 6.15. Queue configuration: Share output



If the protocol **share** is chosen, the controls to configure the output share are revealed, as shown in Figure 6.15. The controls for choosing the share type, along with the file name, are common to those found to set archive, log and input shares.

Production Printing

A number of output options are only available if the production printing licence flag (PRO-PRINT) is enabled. Figure 6.16 shows the production printing options available.

Figure 6.16. Queue configuration: Production printing

Production Print Settings

General

Page scaling: 100%

Imposition

Apply imposition:

Media size:

Duplex mode:

Page scaling: 100%

Separate output queue?

Section cover sheets?

Job Ticket/Job Header

Enable?

Type:

Output Batching

Enable?

Output queue:

Batch the output:

Concatenate jobs:

Temp files:

Daily release time:

Release threshold:

Document template:

Batch header:

Batch trailer:

General options

Options here apply to all PDF files being passed through the queue. The page scaling option allows the size of the PDF contents to be scaled down prior to printing. This can be used to reveal space for additional printed items such as barcodes and text relating to the production process.

Imposition options

Imposition options allow the production user to have A4 or Letter "viewer ready" PDF files laid out on A3 so that they are printed in the most economical way using A3 production print equipment.

This important feature allows a customer to generate and archive PDF files which are suitable for local printing, electronic delivery and viewing, but which can be printed using the best method at the print production house.

When imposition is applied, the DataGateway output filters create a new copy of the document, with the media sizes, simplex/duplex settings and page order re-configured for the new output. An incoming document containing 10 A4 pages may be output as 2 A4 pages,

followed by a single A3 page ready for booklet style folding, followed by 4 more A4 pages. It is possible to have the A3 pages output on a separate queue, so that different production equipment can be used to produce the documents.

If imposition is required it must be triggered from settings applied in the Template Manager (**Apply Imposition** set to **Triggered**). To have a section of a document converted to A3, the document should be created or manipulated via the Template Manager, and the relevant media options should be applied to the document. For further details, see Chapter 7.

Media size, duplex and page scaling can all be set for the pages which will be re-ordered. **Media size** refers to the output media size onto which the pages in the input document will be placed. Choosing **Automatic** will cause the imposition system to choose the most suitable output page size. A4 input pages will be laid out on A3 paper. A3 output can be forced, or, if A4 is chosen for **Media size**, then the input A4 pages will be scaled to A5 size, then laid out, 2 to an A4 page side.

Page scaling is likely to be required, as A3 production printers will often have a slightly smaller printing area than 2 A4 pages, due to mechanical constraints on the way the paper is handled in the machine.

The **Separate output queue** option allows for imposed pages to be output separately from the rest of the document. In this case, it is usually necessary to output **section cover sheets** to indicate which document each section belongs to.

Job tickets

Some production printers cannot make use of PostScript MediaType selectors in the PostScript pagedevice. To select media type (and other features) for these printers, it's often necessary to add proprietary extensions to the print data in the form of a job ticket or header. Currently, this feature allows only KDK comments to be added to the PostScript header to define features required on Kodak Digimaster presses.

Note

WML would expect to extend the job ticket feature as necessary, based on customer requirement.

Output batching

The **batch** output protocol is only available if the production print features are licenced. This allows the jobs processed on a queue on the DataGateway to be stored up in an "output batch". This is intended to help the production house to manage the printing of documents which are "drip-fed" from the customer.

The production print staff can "release" jobs manually, or on a scheduled basis to the final output printer.

A common situation would be a printer releasing all documents which have been received up to a 3 PM cut-off time so that they can be printed and mailed that day.

After output batching has been enabled for a queue, it's necessary to choose the output queue to which the jobs will be released. If none is chosen, it will only be possible to down-

load a zipped archive of the output jobs. Jobs can be batched within the WML PDF filter system (before the PPD has been applied) or after it, so that the jobs are in print-ready (PostScript or PCL) format (**Batch the output**). Batches may be concatenated into single file. The pending output batches may be stored on a RAM disk, or within persistent storage, depending on the agreement on data retention with the customer (**Temp files**).

To control the release of documents, it is possible to specify a daily time at which to release documents, or that the release will be manual (**Daily release time**), or following the build-up of a threshold number of jobs in the output batch (**Release threshold**).

To aid the production process, a batch header and trailer document can be printed containing the number of documents, the sequence start and end numbers and a number of other pieces of information. To define how the header and trailer will appear, templates should be set up in the Template Manager and selected in **Batch header** and **Batch trailer**. The template specified in **Document template** is applied to all the pages in the output documents, and allows the document sequence number, the page number, page count, date and time to be applied to each page in the output documents.

For further details about the output batching functionality, see Chapter 17.

Queue info

The queue info tab (Figure 6.17) displays a summary of queue and printer details for the current queue.

Figure 6.17. Queue configuration: Queue Info tab

Print Platform | File Manager | System Settings | Logout

Queues | Queue Config | Jobs | Platform Settings

Queue configuration for mainPrinter

Switch to queue: mainPrinter | Print status page

Input | Text | Failover | Alerts | PPD Driver | PPD Options | Kyo Settings | Output | Queue Info

Queue Summary

Queue Name	Make & Model	Printer Info	Printer Location	Printer Device URI
mainPrinter	HP 910, hpcups 3.9.12	Main queue for Prescribe data processing		lpd://192.168.0.43/passthru

[View full job listing](#)

Processing Jobs

No jobs are currently processing on this queue

Completed Jobs

No jobs have completed on this queue

6.3. Queue configuration: Example workflow

This section describes the basic steps required to create a new, usable print queue on the DataGateway. It doesn't describe every setting (refer to Section 6.2 for that), but instead gives an overview of the initial configuration of a new queue.

First go to **Print Platform > Queues**. To create a new queue, we will fill in the details in the “Add a new queue” row of the “Print Queues” table.

Figure 6.18. Setting the new queue name



Figure 6.18: Enter the name of your queue in the “Name field”. Use upper and lower case letters, numbers, the dash character and the underscore character. Queue names may not contain spaces.

Figure 6.19. Setting the input TCP port



Figure 6.19: Enter a TCP/IP port number for raw port printing. This is the port on which the queue will accept print jobs. A common port for print applications is 9100 which is the default on the DataGateway. If you will not use raw port printing, simply leave this set to 9100.

Figure 6.20. Setting the queue information



Figure 6.20: Enter a description of the queue here which will help you administer the system.

Figure 6.21. Configuring the output settings



Figure 6.21: Choose an output data transfer protocol and specify the destination address and queue or port. If you are sending data to a printer, the most likely settings here are “LPD”, the IP address or network name of the printer, and “passthru” (usually, a network printer will respond on any LPD queue, so “passthru” is simply a common choice).

Figure 6.22. Adding the new queue



Figure 6.22: Finally, press the “add” button to create the queue.

Figure 6.23. The new queue is ready

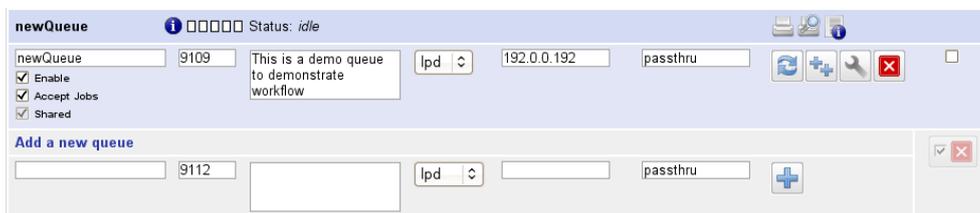


Figure 6.23: The new queue will appear directly above the “Add a new Queue” row. It will be created in the “accepting” and “enabled” states (see the checkboxes to the left of the new queue row), meaning that it will accept new jobs for processing and it will process any accepted jobs immediately.

Figure 6.24. Enter the queue configuration for the new queue



Figure 6.24: Press the “configure” button to enter the queue configuration page for the new queue.

Figure 6.25. Configuring the output PPD driver

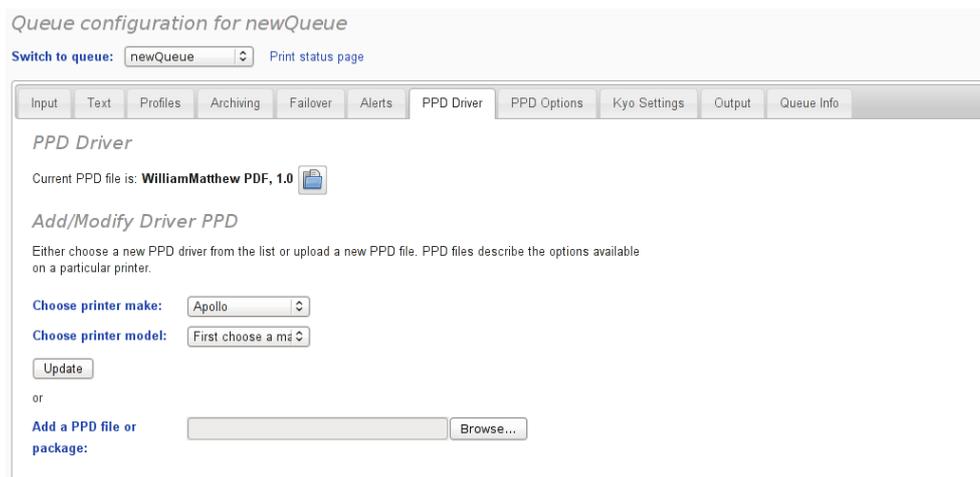


Figure 6.25: The most important configuration setting for the queue is the PPD driver. This is where you define how the DataGateway should format the document before it is sent to

its destination. The PPD driver selects which of the DataGateway's internal printer driver filters is used to convert the job from the DataGateway's internal format (PDF) into printer-ready output.

Refer to the section called “PPD driver” if you need to upload your printer manufacturer's own PPD file. Here, we will assume that the PPD file for your printer has already been uploaded to the DataGateway.

Figure 6.26. Selecting a printer make

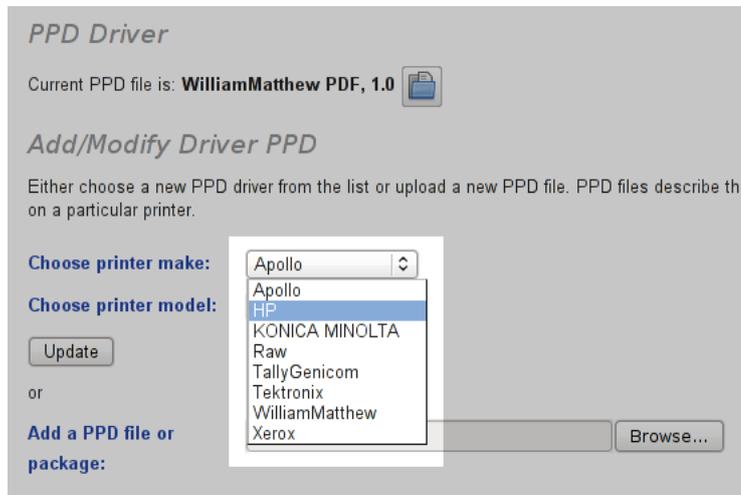


Figure 6.26: Choose the printer make from the drop-down menu. In this example, we're selecting an HP printer.

Figure 6.27. Selecting a printer model

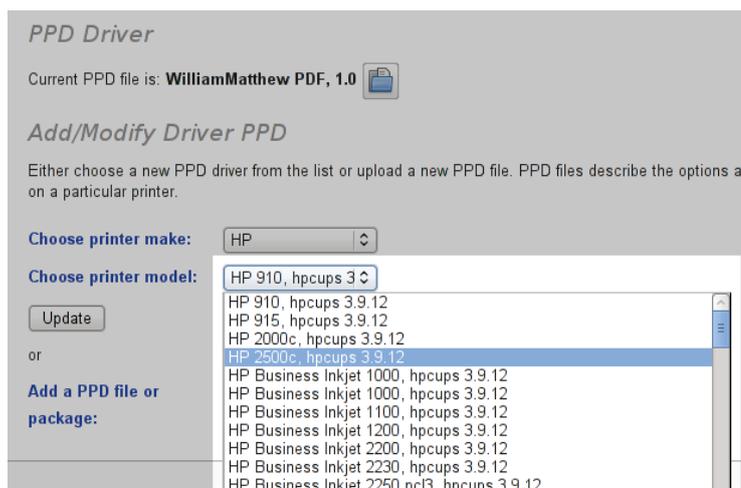


Figure 6.27: Choose the printer model from the second drop-down menu (which will now contain all the HP printers for which which the DataGateway has an installed PPD file). Here, an HP 2500c is selected.

Figure 6.28. Comitting the changes

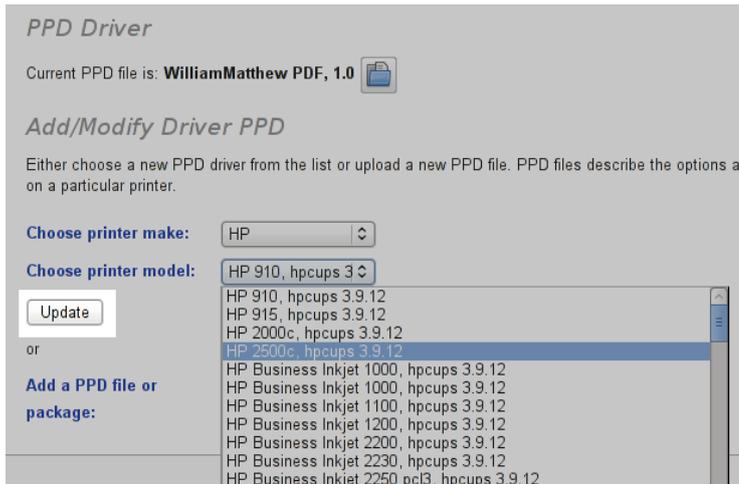


Figure 6.28: Press “Update” to change the the newly selected PPD driver. The “current PPD file” should change to the new value (all new queues have “William Matthew PDF, 1.0” by default).

Figure 6.29. Configuring PPD options

The screenshot shows a configuration window with the following settings:

Section	Option	Value
General	Media Size	Letter 8.5x11in
	Media Size	Index Card 3x5in
	Double-Sided Printing	Off
	Media Source	Auto-Select
	Media Type	Plain Paper
	Print Quality	Normal Color
Installable Options	Duplexer Installed	Not Installed

Update

Figure 6.29: Any printer options available for the selected printer will now be available in the “PPD Options” tab. Configure these to your requirements.

The new print queue is now ready for use. You can send data to the queue using LPD, RAW and IPP print protocols. If you need to use a hot folder for transferring print data to the DataGateway, archive your print data or configure other aspects of the queue, please refer to the sections earlier in this chapter for details.

6.4. Jobs

The **Print Platform > Jobs** page shows lists of currently processing and completed jobs. It shows details for a limited number of jobs; by default the last 500 print jobs are listed. This number can be adjusted in **Print Platform > Platform Settings** by going to the **Cups Control** tab and setting the “MaxJobs” parameter.

Jobs printed on individual queues can be shown using the “Print queues” menu. The size of the job lists can be limited with the “Number of jobs” menu.

If job data is held, in addition to job info (by setting the parameter `PreserveJobFiles` in the **Edit cupsd.conf** tab of **Print Platform > Platform Settings**), then jobs can be re-printed via this interface. Note that re-printing here will be from the original *input data*, so if any meta-data has been applied to the job previously (such as the current date or time) or if a PDF form was applied and has since changed, the print job may render differently from the

originally printed version. For accurate re-printing of jobs, the job archiving and re-print feature should be used (**File Manager > Archive Files**).

6.5. Platform settings

Print settings which are not specific to a single queue are configured via the “Platform Settings” page.

Cups Control

All WML Print Platform products use CUPS for print data transport. This tab allows certain CUPS parameters to be changed by the user, and allows the CUPS daemon to be re-started.

- **LogLevel:** Determines which log messages CUPS output to the system log.
- **MaxJobs:** The number of jobs for which CUPS holds information.
- **Port:** The TCP port on which CUPS listens.

Note

Leave the LogLevel at “notice” if print job reporting features are being used.

Failover

Check “Enable failover mode” if the main print destinations have failed and jobs should be re-routed to the failover queues.

Edit cupsd.conf

This feature is intended for support staff only. This is the main CUPS configuration file.

Edit mime.types

This feature is intended for support staff only. This is the configuration file which determines how the file type of input data files is identified.

Edit mime.convs

This feature is intended for support staff only. This is the configuration file which determines how data files are converted into the DataGateway's internal format, PDF.

Chapter 7. Template Management

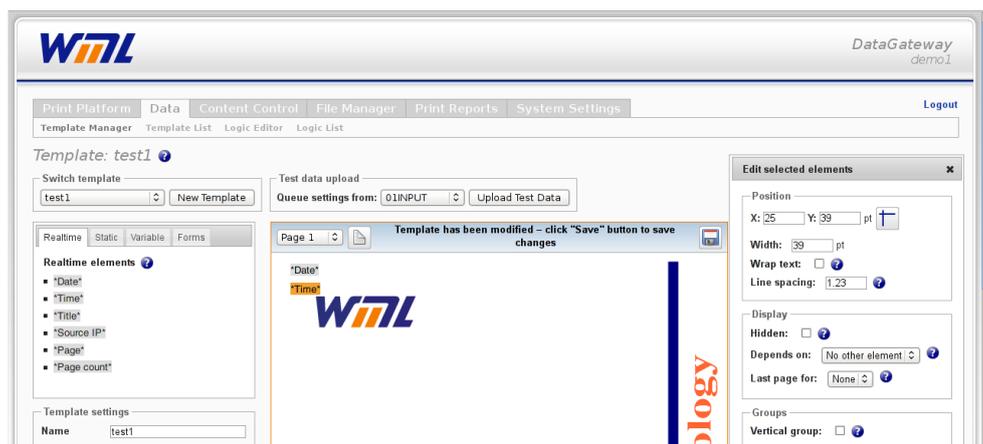
7.1. Introduction

The DataGateway's main tool for data manipulation and document creation is the *template manager* (Figure 7.1). This WYSIWYG tool allows you to build up a *template* which can be applied to print jobs as they are processed by the DataGateway.

A *template* is a collection of *overlay elements* which are individual items of text or images. The contents of the template are merged with the incoming print data to create the finished print document.

A unique feature of the WML DataGateway is that all data manipulation is applied natively in the PDF format. This means that the output document is always available as a PDF for archiving or electronic delivery.

Figure 7.1. The browser-based template manager



It's easy to build up a template using the DataGateway's template manager. Simply upload your PDF Forms and drag and drop text elements into position. It's a quick, easy job to build up a template which will enhance the quality of your print output.

Figure 7.2. A finished example template

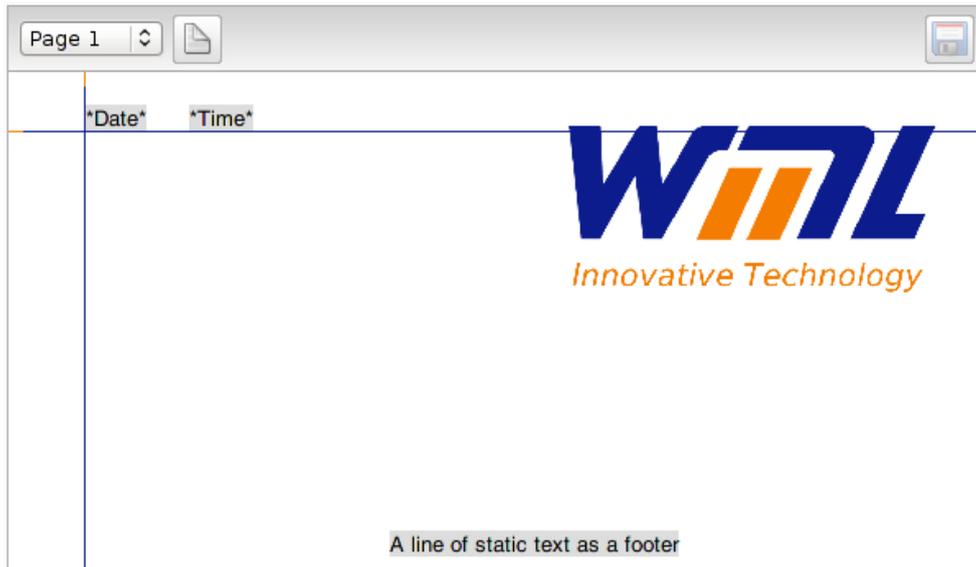


Figure 7.2 shows a finished template containing date and time realtime data elements, a static element, and the same PDF form as shown in Figure 7.5.

Figure 7.3. The result of the example template viewed in the Adobe PDF Reader



Figure 7.3 shows the result of passing a simple input print job to be processed by the example template.

7.2. Overlay elements

An **overlay element** is an item of data to be applied to a page of a print job. This might be a piece of text, a group of texts, an image or even a complete PDF. In fact, these are the four categories of overlay elements; **text elements**, **data group elements**, **image elements** and **PDF form elements**.

Each overlay element has an **element type**. The different types are:

- Realtime data: A text element whose value is generated as the data is processed (e.g. the date).
- Static data: A text element which always has the same value.
- Variable data: A text element which is defined with a key. The value is provided at processing time with input print data. Variable data elements can be simple, single elements (e.g. “customer_name” filled with the value “Mr Smith”) as well as **Data Groups** to lay out tables and multiple paragraphs.
- Data group: A set of text elements which should be laid out over a given area. Used for documents with variable size, such as those with a variable number of paragraphs, or variable length tables.
- PDF form: A PDF file containing one or more pages.
- Barcode: An image element, generated by converting a piece of text within the input print data into a barcode.

An element can be added to a single page of a document (perhaps page 1 or page 3) or it can be added to all pages of a document.

Figure 7.4. Three text elements; a date, a time and a piece of static text



Text elements have position and font specifications, controlled by the “Element editor” (seen in Figure 7.1 and Figure 7.7). Image elements are simply given a position. PDF form elements are laid onto the page, as if they were a transparency on a projector.

Figure 7.4 shows three text elements, positioned towards the top left of page 1 of the template. The snap-to lines are visible and the static text element is highlighted in orange, showing that it is the currently selected overlay element.

Figure 7.5 shows a single PDF form element containing a WML logo, which is active for page 1 of this template.

PDF form elements may have multiple pages. If you add a multi-page PDF form to page 1 of a template, then page 2 of the form will print on page 2 of the template, page 3 of the form on page 3 of the template and so on, as long as there are enough pages in the template or incoming print job.

Figure 7.5. The logo seen here is a PDF form element, applied from an uploaded PDF file



7.3. Overlay templates

An *overlay template* combines text elements, image elements and PDF forms which can be applied to individual pages, or to every page in the input print job.

It also includes a number of *template settings* such as the template name, the number of pages for the template and the page size.

An overlay template is stored in a single, XML-based configuration file on the DataGateway. Overlay templates can be created on one DataGateway and then exported and copied onto other DataGateways, either individually using the Template Manager, or as a group using the facilities for backing up the DataGateway configuration in **System > Settings**.

An overlay template can be “bound” to a queue, so that all data passing through that queue will be processed by the overlay template. The overlay template is bound to the queue in the queue configuration, see Chapter 6.

Alternatively, the WML print filters can select an overlay template based on the existence of text triggers within the input print job. For example, an overlay template named “CompanyInvoice” might be triggered when a document containing the text “INVOICE” were processed. For more information about logic-based triggering, see Chapter 8.

7.4. Overlay template list

To create the first overlay template, or to view a list of templates, select the menu item **Data > Template List**.

Figure 7.6 shows an example of the template list. It is a table with one line for each overlay template. The template name is shown, as well as the *document basis*, the number of overlay elements in the template and a column of action buttons.

Note

The meaning of *document basis* will be described later.

Figure 7.6. The overlay template list



To create a new overlay template, enter a name for the template. The name should be unique because it is used as an identifier for selecting the active overlay template (either manually or by means of a trigger in the input data). Click on the add button (+) to create the new template.

You may also duplicate an existing overlay template, by clicking on the duplicate button (⌘) for the template that you wish to copy. A new template will be created and added to the list.

While a newly created overlay template does not contain any overlay elements, a template created by duplication will contain copies of the elements in the duplicated template.

You can change the name of an existing overlay template by editing the value in the text field and clicking on the update button (↻).

Note

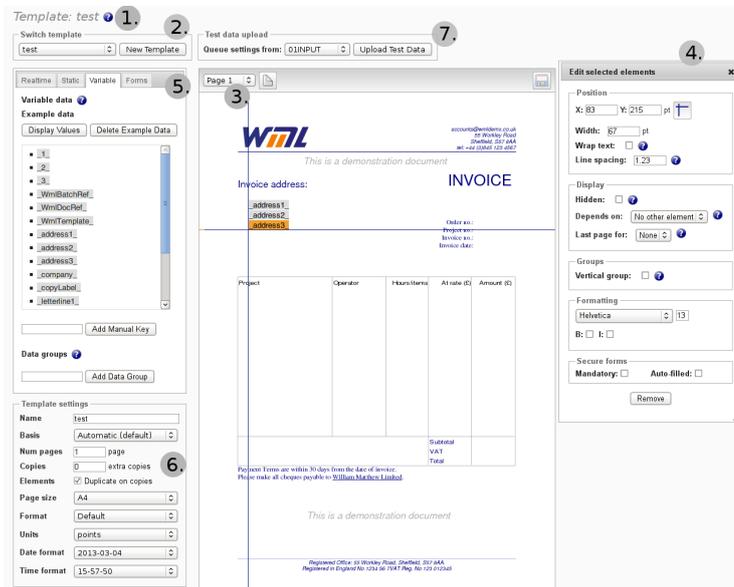
If you modify the name of a template which has been assigned to a particular queue, you will need to re-assign the template to the queue using the new name.

From the template list, it is possible to “drill down” into the template manager for any of the listed templates. Edit a template by clicking on its configure button (⚙).

7.5. Overlay template manager

The overlay template manager was shown previously in Figure 7.1 and is presented in full view below (Figure 7.7).

Figure 7.7. The overlay template manager



The template manager consists of the following components:

1. The title - this tells you the name of the current overlay template.
2. A top toolbar which allows you to switch to a different template, or create a new one.
3. The template “canvas”, showing you how the overlay elements will print. At the top of the canvas is the “canvas toolbar”.
4. The element editor (to the right of the canvas) which contains settings for the currently selected overlay element(s).
5. The overlay elements box, from which you can drag text elements onto the canvas, and activate PDF form elements.
6. The template settings box, which allows you to change the overlay template's attributes.
7. The Test Data box, which contains an upload field which you can use to upload an example print job. This is used to test your overlay template without the need to fully set up a print queue.
8. The template download box (not visible in the image; below the Template settings box). Allows the user to save an archive of the current template.

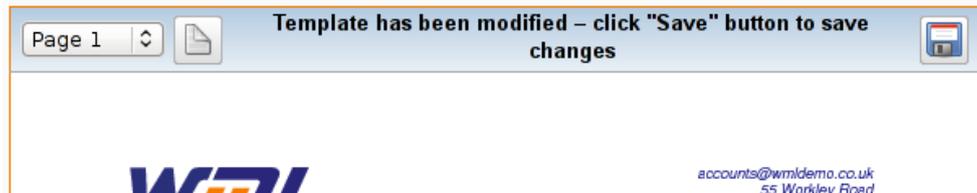
Saving changes to the template

The template manager is a browser-based front-end to the DataGateway appliance. It sends messages to a back-end program on the DataGateway which writes changes to a configuration file.

In general, changes made in the left column are immediately saved on the DataGateway without any further action. In contrast, changes to the position, width, font size and weight

of any text elements are *not* immediately saved. Any such changes will cause a “modification warning” to display at the top of the canvas. Figure 7.8 shows an example of this warning. To commit changes to the DataGateway it is necessary to press the save button (📁).

Figure 7.8. This template has been modified



Template management

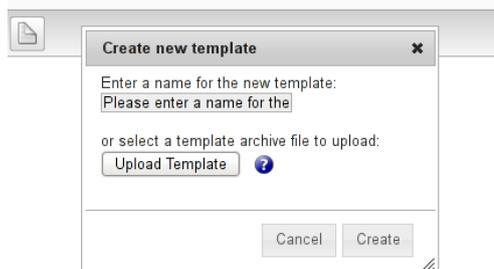
Switch template

Immediately below the title are two controls: A menu which allows you to switch to a different template and a button to create a new template.

If you try to switch to a different template and you have made changes to text elements which have not yet been saved, the DataGateway will open a dialog to ask if you are sure you wish to switch and lose your un-saved changes.

If you click the “Create template” button, a dialog will open asking you for a name for the new template (Figure 7.9). You can also choose to create a new template by uploading a template archive (as saved under **Template operations**).

Figure 7.9. The new template dialog box



Template operations

Below the **Template settings** box is the **Template operations** box. This allows you to duplicate the template or save the template to a file.

Pressing **Duplicate** will create a copy of the current template with an automatically generated name (the name of the original template with “copy” appended).

To save the template, press **Create Archive**. This makes up an archive file containing the layout of the template and all PDF forms which are part of it, and allows you to download it with your browser. The file is called **template.tar**.

Note

To upload the template.tar file to another DataGateway, so that the template can be used on that system, use the **New Template** button in the **Switch template** box. The **Create new template** pop-up has an **Upload Template** button.

Template settings

The template settings box provides a number of settings which are “template-wide”. That is, they affect all elements in the template, or they affect the way the template operates. Figure 7.10 shows the template settings box.

Note

Any change made in any of the fields in the template settings box are *immediately* saved on the DataGateway.

Figure 7.10. The template settings box

The image shows a 'Template settings' dialog box with the following fields and values:

- Name:** Formatted Text
- Basis:** Automatic (default)
- Num pages:** 1 page
- Copies:** 0 extra copies
- Elements:** Duplicate on copies
- Page size:** A4
- Units:** points
- Date format:** 2010/09/24
- Time format:** 3:36 PM

Name

This is simply the name of the template. The template name is modified by changing the value here. (It is also possible to change the name in the template list page, see Section 7.4).

The template name may contain letters, numbers, spaces, dashes and underscores.

Note

If you modify the name of a template which has been assigned to a particular queue, you will need to re-assign the template to the queue using the new name.

Basis

The *Document Basis* determines the way in which the input print document and the template, with any PDF forms, are merged together into a finished document.

The PDF document format has a feature which allows new objects (such as lines of text or images) to be added to the pages of the document simply by appending data to the end of the file. The DataGateway makes use of this feature.

The PDF to which all other objects are appended is the “document basis”.

Note

The number of pages in the completed document will be the same as in the document basis.

The Basis is set to one of four possible values, which are described below.

Basis: Input print document

If the input print document is used as the basis, then the output print jobs will always have the same number of pages as the input print document.

This mode is used when you expect print jobs of different sizes to print on your queue.

Note

Use input print document mode if you want to apply the same template elements to the first few pages of every document, but your print jobs are of variable length (in pages).

Basis: Manual number of pages

If you specify “Basis: Manual number of pages”, then the DataGateway will create an empty PDF containing the specified number of pages and use that as the basis for building the output document.

In this mode, output print documents will *always* contain the number of pages specified in the “Num pages” field in the template settings box.

Note

Use this mode if your input print documents contain fewer pages than you would like to print.

For example, you could add terms and conditions to the second page of your single-page invoice documents by setting Basis: Manual number of pages; Num pages: 2 and placing a PDF form containing the terms on to page 2 of your template.

Basis: Automatic

This is the default basis for new templates.

In “Basis: Automatic mode”, the WML print filters will choose between “Basis: Input print document” mode and “Basis: Manual number of pages” mode. It makes this decision based on the number of pages in the input print document, and the number of pages in the template.

If the input print document contains the highest, or equal highest number of pages, the input print document is used as the basis for building the output document. Otherwise, the filter creates an empty PDF containing the number of pages specified for the template and uses that as basis.

Basis: PDF form

In certain cases, it is desirable to use one of the PDF Forms in the template as the document basis.

Each active PDF form *on page 1* of your template will appear as an option in the Basis menu Like this:

PDF form: myfile.pdf

For example, you might add a multi-page PDF to your template and use this as the Basis. The advantage is efficiency. It is less processor-intensive to use the largest PDF form as the basis when compared with the “Basis: Manual number of pages” mode.

Note

When you are using the DataGateway to print structured input data (CSV or XML formatted key-value pairs) it is highly recommended that you use a PDF form as your document basis.

Num pages

The number of pages in the template. The exact meaning of this number is dependent on the document basis.

In “Basis: Input print data” and “Basis: Automatic” modes, this is *the number of pages which may be modified by the template*. If Num pages is set to “2”, and a single page document is printed, the elements on the second page of the template will be ignored.

In “Basis: Manual number of pages” mode, Num pages is *the number of pages which the output document will contain*. The output document will always have this number of pages, even if some of the pages are blank.

If a PDF form is chosen as the document basis, then this field will be greyed out and set to the number of pages in the chosen PDF form.

Page size

This is the expected page size of the input print data. You should match this to the size of the documents which you will be printing.

If you do not match the page size of the input print data and the page size specified here in the template, your overlay elements will not print in the correct positions.

Note

If the WML print filters detect that you printed a A4 document through template set to US Letter (or vice versa), it will print a warning page to the effect.

Units

The units in which all overlay element positions will be specified. This defaults to “points” (1/72 of an inch) which is commonly used internally within PDF documents. Other options are *mm*, *cm* and *inches*.

Date format

This is a drop-down menu with a number of date formats such as “24th September 2010” and “2010-09-24”.

The date format will be applied to all Date realtime data elements on the template.

Note

In the current revision of DataGateway, it is not possible to have differing date formats within the same template. For example, you could not have Date realtime data element on page one filled in as “14th Sept 2010” and one on page two filled in as “2010-09-14”.

Time format

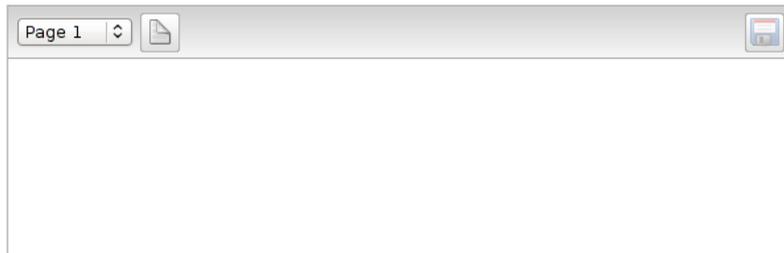
Like the date format menu, this is a drop-down menu with a number of formats for printing the current time in Time realtime data elements.

Note

In the current revision of DataGateway, it is not possible to have differing time formats within the same template.

The canvas

Figure 7.11. The template manager's “canvas”



The canvas is the main work area of the template manager (see Figure 7.11 which shows an empty canvas). It displays the current template page as it would print. Any active PDF form elements are displayed, along with text elements and “snap-to” lines, if they are enabled. You can select one or more elements (they will turn orange when selected) and move them by dragging. Controls in the element editor allow changes to the font as well as enabling you to fine-tune the position and width of the fields.

The canvas toolbar

At the top of the canvas is the *canvas toolbar*. This contains a menu to switch to other pages of the template (the *pages menu*), a media control button and a save button to write changes to the DataGateway.

The pages menu

In Figure 7.11, the canvas of page 1 of the template is displayed. This is seen in the menu at the far left of the canvas toolbar (the *pages menu*). To modify overlay elements on a different page, select the page from the pages menu. The pages menu contains one entry for each page in the template and additional entries for several special pages: “Others”, “Odd”, “Even” and “All pages”.

Note

New pages are added to the template in the template settings box (see “Num pages”).

“Others”

By selecting “Others”, you are placing overlay elements *which should appear on any page of an input document which is not covered by the template*.

For example, imagine your template contains a single page, on which you place a company letterhead. Now imagine you place a simpler PDF form containing only a company logo on the “Others” page. If you print a one page document through the template, the letterhead will be printed on page one. If you print a two page document, the letterhead will print on page one and the logo will print on page two. If you print a one hundred page document, the letterhead will be printed on page one, and on every other page, the logo will be printed.

“Odd/Even”

By selecting “Odd” (“Even”), you are placing overlay elements which should appear on any odd (even) page of the print job.

“All pages”

If you select “All pages” from the pages menu, then any elements you place will appear on every single page printed through the template.

Note

Text elements which are placed on the “All pages” canvas will be shown in blue text. These elements will appear on all other pages of the template. For example, if you switch to page 1 of the template, the elements you placed on the “All pages” canvas will appear alongside those which will only appear on page 1.

Saving changes

To commit changes press the save button ()

The element editor

The element editor contains fields controlling the position, width, font-face, font-weight and font-size of selected elements.

Position and width

The X and Y fields show the current X and Y position in the current units, with respect to the top-left of the page. These values can be edited and the selected element will move to the new position.

Note

The units default to “points”. The units can be changed in the template settings (see “Units”).

The *Width* field shows the element width in PDF units (points). The width is the required width of the element when printed. It is used in the WML Secure Forms module to size the editable field for each variable data text element. You can also check the *Wrap text* option to wrap the element text to the specified width. The *Line spacing* field governs the spacing within a word-wrapped text element. The line spacing is specified as a proportion of the font-size.

Formatting

The font face can be selected from the drop down menu. Directly to the right of the font menu is a field to specify the font size.

The available fonts are: Times New Roman, Helvetica, Courier, Symbol and Zapf Dingbats. These are the basic “built-in” fonts which are available in every PDF document.

The font drop down menu also provides options for barcoding. If enabled, you can also choose an option to convert a text element to a barcode image.

The current font can be emboldened or italicized by checking either the *B* or the *I* boxes, respectively.

Deleting elements

To delete selected elements, press the “Remove” button, or press the delete key on your keyboard. As with other changes to elements, the changes will not be saved on the Data-Gateway until the save button is pressed.

Snap-to lines

The template manager provides “snap-to” guidelines which can be enabled and disabled as required. To enable the snap-to lines, press the snap-to lines button (). The snap-to lines can be seen in Figure 7.4.

If you drag an element towards the snap-to lines, the element will dock with the lines, allowing you to line up a set of elements.

To disable the snap-to lines, press the snap-to lines button again.

Working with multiple elements

It's possible to select multiple elements and apply the same font-face and font-size to the set. Multiple selected elements can also be dragged as a group.

You can also modify the X and Y positions for a group; however, if you select multiple elements, and they are offset from each other, then changing the position will maintain the offset between them. The position of all selected elements will change by the amount that the first element moves.

The overlay elements user interface

The overlay elements user interface box, shown in Figure 7.12, is the “source” for all overlay elements. From here you can manage realtime data, static data, variable data and PDF forms. Text elements are dragged onto the canvas. PDF forms are “activated” for the current page of the template.

Note

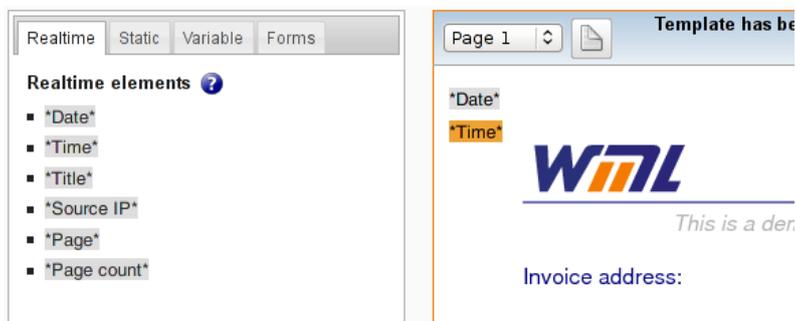
For a description of each overlay element type, please see Section 7.2.

The overlay elements interface consists of four tabs, one for each type of overlay element. Each tab is described here.

Realtime elements

Figure 7.12 shows the realtime data tab. Four realtime data elements are displayed: Date; Time; Title and Source IP. The value of these fields is filled in when the job is processed (i.e. just as it is printed).

Figure 7.12. Overlay elements - Realtime tab



To place any of these elements on the canvas, drag the element from the overlay elements realtime tab and drop them in the desired location.

Figure 7.12 shows a date and a time element placed on the canvas (the time element is selected and is highlighted orange). Note that realtime data elements are enclosed by '*' characters. The date element appears on the canvas as "*Date*" the Time element as "*Time*".

The date and time realtime data elements are intended to be used in time-stamps to prove the date at which a job was processed for printing or archival. The format in which the date and time are printed is configured in the template settings.

The title realtime data element is filled in with the job title. This may be set by the user if the print data has been transferred to the DataGateway using the LPD or IPP protocols, which allow the specification of a job title. If a hotfolder has been used to transfer the print data onto the DataGateway then the filename of the print file in the hotfolder will be used. If raw TCP/IP data transfer has been used, or if the title has not been set when using LPD or IPP; the title will print as "unknown".

The "Source IP" realtime data element will contain the IP address of the sending computer system if either the LPD or the raw TCP/IP protocols have been used to transfer print data onto the DataGateway.

Note

Unfortunately, the sender's source IP address cannot be determined when the IPP protocol or hotfolders have been used for print data transfer.

Static data elements

Static data is the simplest overlay element. It is an unchanging piece of text (a word or phrase) which always prints in the same way. Static data can be used to make simple amendments to documents without the need to go back to the original artwork editor.

Figure 7.13. Overlay elements - Static Data tab



Figure 7.13 shows the static data tab, with a single static data element containing the text “Copyright 2013”. To place the element, drag and drop it in the usual way. Multiple copies of the static element may be placed on the canvas.

To create a new static data element, with different contents, fill in the blank text field and press the button “Add Text”. The new element will appear below “Copyright 2013”. You can now drag this element onto the canvas.

Note

When a template is accessed, the static data tab will show only static data elements which are placed on the template. If you add a static element with the “Add Text” button, then don't use it, it will not be available the next time you load the template in the template manager.

Variable data elements

A variable data element is a text element with a “key”. When CSV or XML input print data is received, specifying the value of this key, the value is printed at the location specified by the element and in the specified font.

The use of variable data input - also referred to as “structured data” - is covered in greater length in Section 7.7: “Using structured input data”.

Figure 7.14. Overlay elements - Variable Data tab, initial empty view

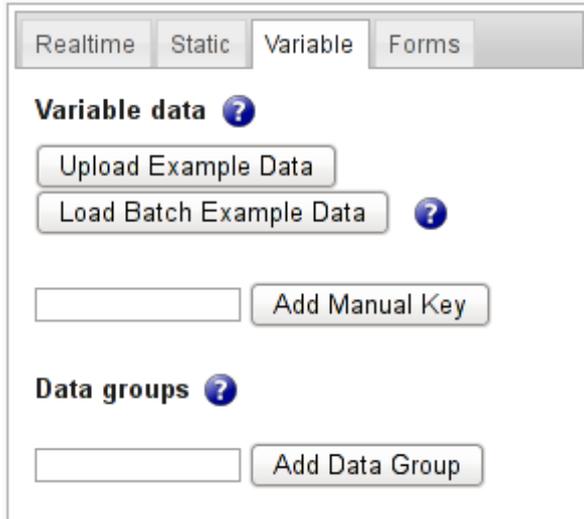


Figure 7.14 shows the variable data tab in its initial, empty state. There is a button to “Upload Example Data” and a field to add a manual key. These are the two methods you can use to create a variable data element. First, we’ll cover manually specified keys.

Figure 7.15. Overlay elements - Variable Data tab after adding a manual key

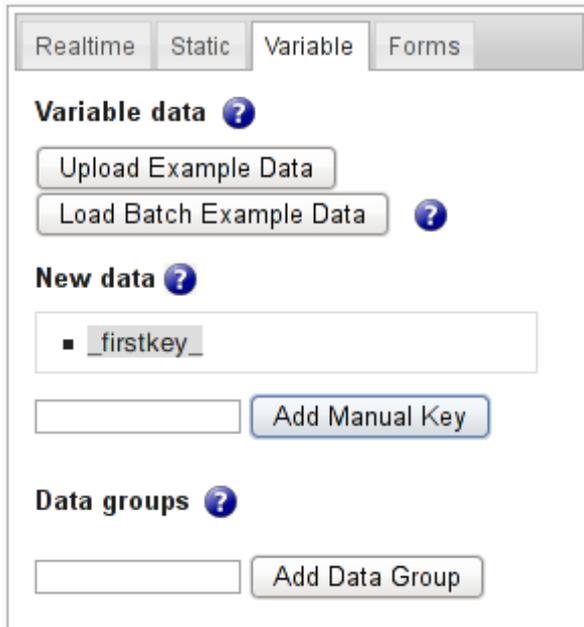
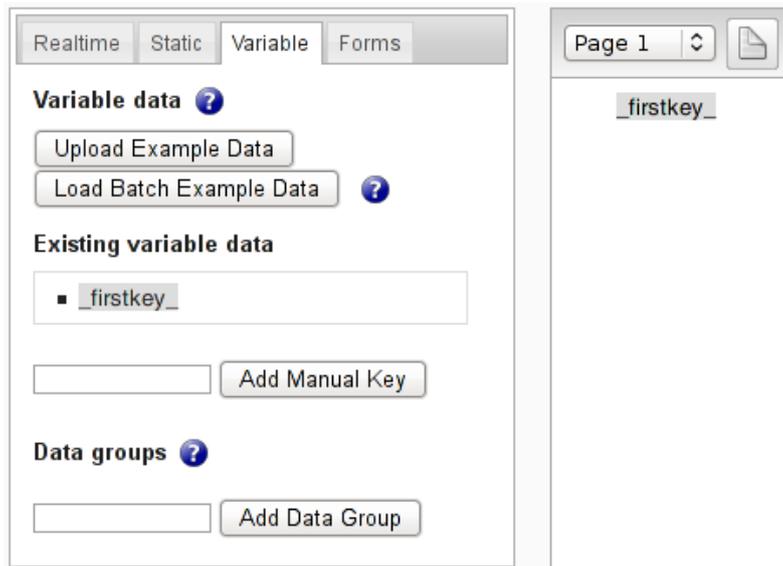


Figure 7.15 shows the result of entering “firstkey” into the manual key text field and pressing “Add Manual Key”. A heading “New data” has appeared with a single item “_firstkey_”. Note the underscore characters enclosing the key. These indicate that this is a variable data element and are not part of the key. They distinguish variable data elements from realtime data elements, which are enclosed by '*' characters and from data group elements, enclosed by '/' characters.

Figure 7.16. Overlay elements - Variable Data tab showing a saved key



This variable data element has not yet been placed on the template canvas. Figure 7.16 shows the variable data tab after the key has been placed. “firstkey” is now in a (short) list under the heading “Existing variable data”.

Manually specifying keys is not the most convenient way to add variable data elements to an overlay template. A more efficient method is to upload an *example data file*. This is a structured data file, arranged in CSV format or WML's simple key-value XML format. The template manager processes this file and extracts the keys, and the example values from the file.

Example 7.1. An example variable data file in CSV format

```
name,Mr Jones
address,24 Tenants Drive
```

Example 7.1 shows a very simple example data file containing two key-value pairs. This file is formatted as comma separated values. The first column contains the keys - in this case “name” and “address” - and the second column contains the corresponding values - “Mr Jones” and “24 Tenants Drive”.

Note

The example variable data file should be exactly the same as the file which would be sent to print. It is simply an example structured data print file.

Figure 7.17. Overlay elements - Variable Data tab showing keys

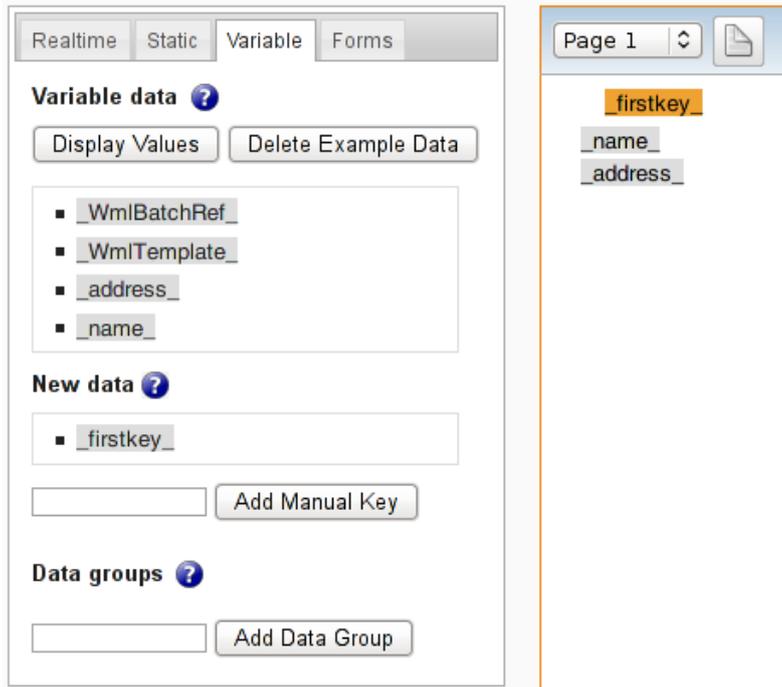
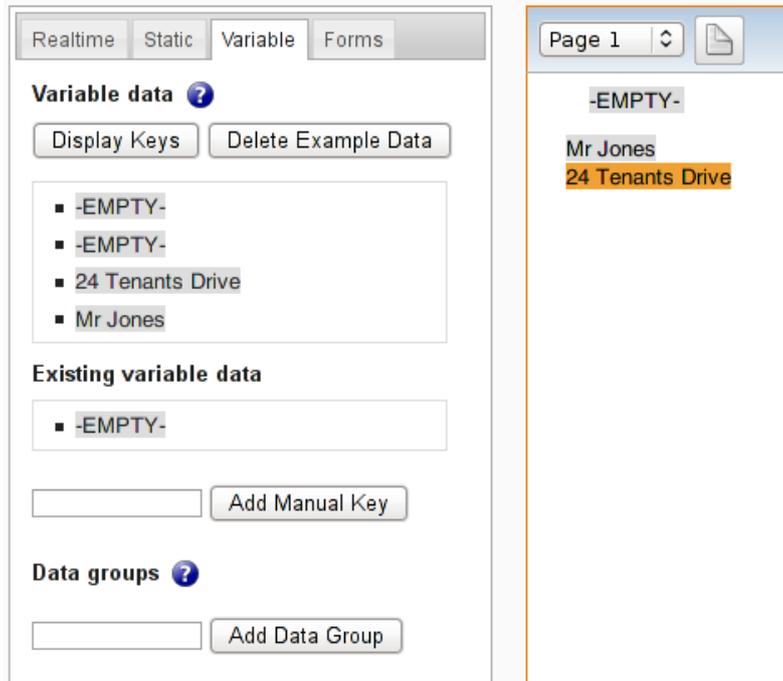


Figure 7.17 shows the result of uploading the example shown in Example 7.1. The “Upload Example Data” button has been replaced by two new buttons and a list of the keys in the uploaded example data (“address” and “name”).

Figure 7.17 also shows firstkey, name and address placed on the template canvas. In common with all other text elements, multiple copies of each variable data element can be placed on the canvas. Here, the “name” key could be printed in several different locations on the same page.

The new buttons are “Display Values” and “Delete Example Data”. “Delete Example Data” simply clears the example data from the template, leaving only any existing variable data elements (i.e. those which have been placed on the canvas and saved).

Figure 7.18. Overlay elements - Variable Data tab showing values



“Display Values” will change the view in the template canvas to show the example value of each variable data element. This is shown in Figure 7.18. Showing the example values provides a better impression of the way the document will print.

Note

The manually created key “firstkey” has no example value and so it is displayed as “-EMPTY-”.

Data groups

Variable data may be brought together in *data groups*. A data group is an element which allows you to control the display of a sequence of related data, with the placement of each data member being determined automatically.

When the input data is processed, the elements assigned to a data group are positioned in sequence on the page, according to the size and spacing attributes given for the data group. If the elements fill up the space available for the group then a new page is created, and subsequent elements are placed there. This allows the flexible layout of variable length documents such as letters and invoices.

Figure 7.19. Overlay elements - Adding a data group



Figure 7.19 shows the result of entering “datagroup1” into the data group text field and pressing "Add Data Group". A heading “New data groups” has appeared with a single item “/datagroup1/”. Note the slash characters enclosing the data group name. These indicate that this is a data group element and are not part of the data group name.

When the data group is placed on the canvas, its name will move to a list under the heading “Existing data groups” (Figure 7.20). Data groups are highlighted in pink on the canvas, with a grid background that indicates the area used by the group. The area can be adjusted using the width and height fields in the element editor. Other attributes, such as font size and line-spacing, can be set in order to control how the elements in the data group will be formatted.

Figure 7.20. Overlay elements - Canvas with data group



The first data group added to the canvas is the “master” data group. You will be able to specify a set of variable data elements to be included in the group, and to control the layout of the group. Subsequent copies of the data group will define “continuation” areas, into which elements defined for the master group may overflow. These can be on the same page or on a special *continuation page*.

Note

It is only possible to edit the variable data keys for a group through the master group element. You can adjust the layout and formatting of a continuation group, but the list of variable data will be disabled.

Some examples of using data groups to display variable input data can be found in Section 7.8.

Forms

Figure 7.21 shows the PDF forms tab. This interface has a menu to select and activate PDF forms which have already been uploaded to the DataGateway, a button to upload new PDF forms to the appliance and a list of active forms. In this example, no forms are active for the template.

Figure 7.21. Overlay elements - Forms tab with no active forms

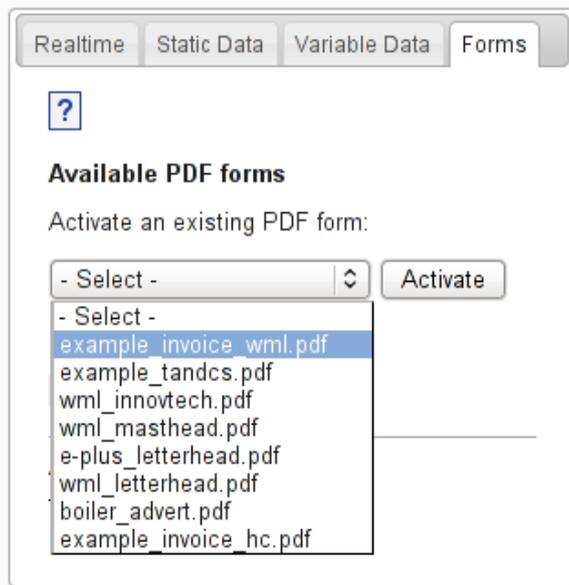
The screenshot shows a web interface with a tabbed menu at the top containing 'Realtime', 'Static Data', 'Variable Data', and 'Forms'. The 'Forms' tab is selected. Below the menu is a blue question mark icon. The main content area is divided into two sections. The first section, titled 'Available PDF forms', contains the text 'Activate an existing PDF form:' followed by a dropdown menu showing '- Select -' and an 'Activate' button. Below this is the text 'or upload a new PDF form file:' followed by an 'Upload Form File' button. The second section, titled 'Active PDF forms', contains the text 'There are no active PDF forms'.

Any PDF forms which you wish to use with the DataGateway must first be uploaded to the appliance. Click on “Upload Form File” and browse to your PDF. The file will be uploaded to the DataGateway and processed by the internal software to ensure it can be successfully parsed.

Note

If the PDF cannot be parsed, perhaps because it contains a feature unsupported by the WML PDF software, then an error message will display. Please contact your vendor with details of the problem, ideally sending them a copy of the problem PDF.

Figure 7.22. Overlay elements - selecting a PDF form to activate



Once you have uploaded your PDF forms, select a form to activate. Choose the form in the “Activate an existing form” menu and press the “Activate” button. An image of the first page of the PDF form will appear on the canvas and box will appear in the “Active PDF forms” list. This will appear as shown in Figure 7.23.

Figure 7.23. Overlay elements - Forms tab with one active form

The screenshot shows a web interface for managing PDF forms. At the top, there are four tabs: 'Realtime', 'Static', 'Variable', and 'Forms'. The 'Forms' tab is selected. Below the tabs, there is a help icon (a question mark in a blue box). The main content is divided into two sections: 'Available PDF forms' and 'Active PDF forms'. In the 'Available PDF forms' section, there is a text prompt 'Activate an existing PDF form:', followed by a dropdown menu showing '- Select -' and an 'Activate' button. Below this is another text prompt 'or upload a new PDF form file:' followed by an 'Upload Form File' button. The 'Active PDF forms' section contains a light blue highlighted box with the following information: 'Filename: example_invoice_wml.pdf' and 'Start on: Page 1' with a 'Del' button.

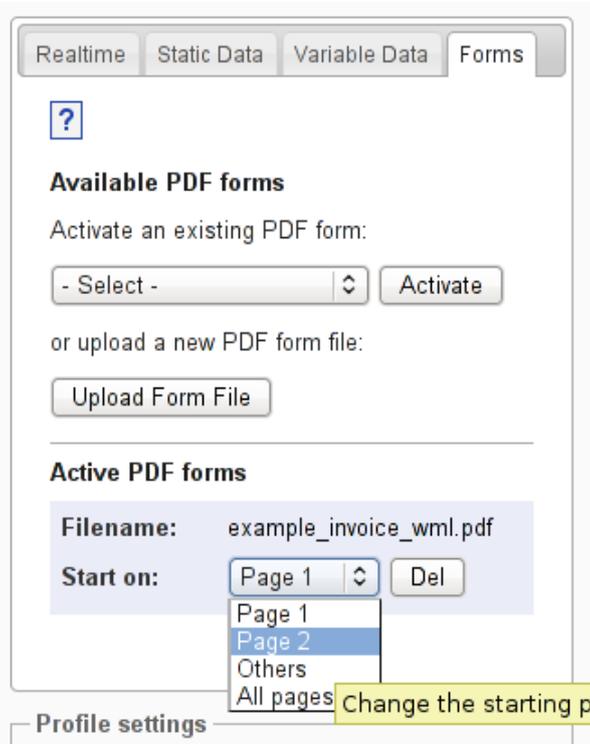
The PDF form will be activated on the currently selected page of the template. The first page of the form will appear on the currently selected template page. If the PDF form should appear on all pages, then change to the “All pages” template page, and then activate the form there.

It's possible to move an active PDF form from one page to another. Each form has a menu listing all the pages in the template. Simply choose a new page for the form (Figure 7.24). To delete the form, press the “Del” button and it will fade out.

Note

All the actions relating to PDF forms are immediately saved on the DataGateway.

Figure 7.24. Overlay elements - Changing the start page for a PDF form



Testing print data

Once you have created your template you will want to verify that it produces good output. The **Test data upload** interface provides the facility to upload an example print file via the web interface. The test data file is processed as if it were a print job and a PDF is returned to your browser. This allows you to test your template without the need to set up a print environment.

Figure 7.25. Test data upload box

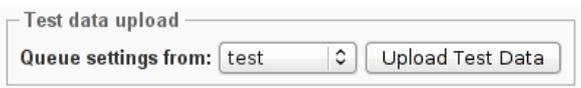


Figure 7.25 shows the test data upload box. It consists of a menu to choose the queue whose settings should be applied to the test, and an upload button.

You can upload any data which can be processed by the DataGateway. The print data format will be automatically identified and the data will be passed through the relevant WML data filters to apply the template. If you have a browser plugin to view PDF files, the resulting file will appear in a new browser tab or window.

Note

If your browser blocks pop-up windows, then you will need to allow pop-ups for the DataGateway.

It's necessary to choose a queue because some print jobs can only be processed with the help of per-queue settings. An example is CSV or text data. There's no easy way for a machine to distinguish between comma separated values and plain, formatted text, so one of the per-queue settings is an option to tell the queue to treat text data as CSV or formatted text. If you are testing CSV print data, then ensure you test with a queue which is set to treat text data as CSV.

Other per-queue settings which may affect the output from your template include Prescribe emulation settings and formatted text emulation settings.

Warning

Some care should be taken when choosing the **Queue settings from** option. If you choose a queue which applies document content control (DCC), you may see a status PDF from the DCC system, instead of the expected PDF generated from your template. This is because DCC typically outputs onto a separate **DCC output queue**.

If the print file is accepted for processing, but there is an error in your template, or a problem with the queue you are using for the test, you may see a WML error page PDF appear in your browser. This will be a plain document stating that an error occurred while manipulating the PDF. The message in the error may indicate a solution to the problem. If not, please contact your vendor for assistance, with a description of the error.

7.6. Overlay Templates: Example workflow

This section steps you through the basic work-flow of creating and applying a template. This will not cover every feature of overlay templates, but provides a starting point for a new user and an overview of the process.

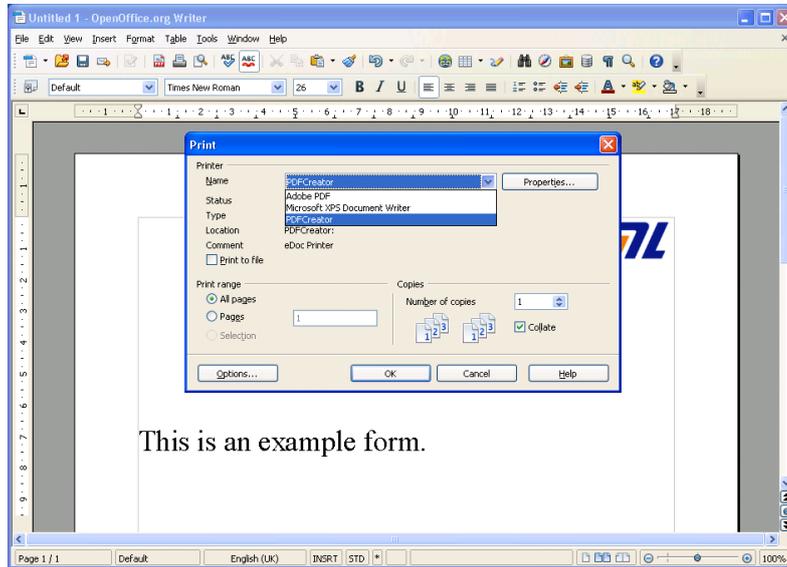
Create a print queue

If you do not already have a print queue configured, follow the instructions in Section 6.3 (Chapter 6) to create a queue. In this workflow example, we'll assume it's called "newQueue" (this comes into play later on when we test the template).

Create PDF forms

You can use almost any "artwork" program you like to generate your PDF forms. You could choose a word processor such as OpenOffice.org or Microsoft Office or a vector graphics program such as Adobe Illustrator or the open source Inkscape. If the program does not have its own PDF output function, then a PDF printer driver can be used to generate PDF output by "printing to a file".

Figure 7.26. Printing to PDF with OpenOffice.org and PdfCreator



Examples of PDF printer drivers which can be used include Adobe's Distiller and the open source PdfCreator (Figure 7.26, see www.pdfforge.org/pdfcreator for a download).

Note

The PDF specification is very broad and the DataGateway does not support every feature. If you have trouble during testing of your PDF forms, please contact support via your vendor.

Create and name a new template

Figure 7.27. Template manager before any templates have been created



If no templates have been created on your DataGateway, the template manager will be empty, with just a link to the template list page (Figure 7.27).

Figure 7.28. The empty template list

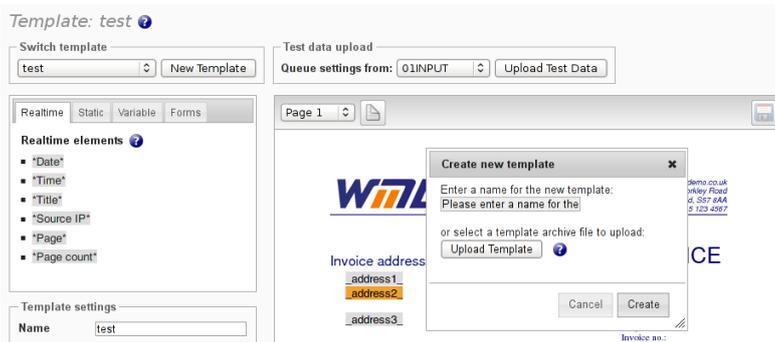


To create a new template from the template list page, enter a new template name in the text field and press the add button (+).

Note

Template names may contain upper and lower case letters, numbers, the dash character and the underscore character.

Figure 7.29. Creating a new template within the template manager

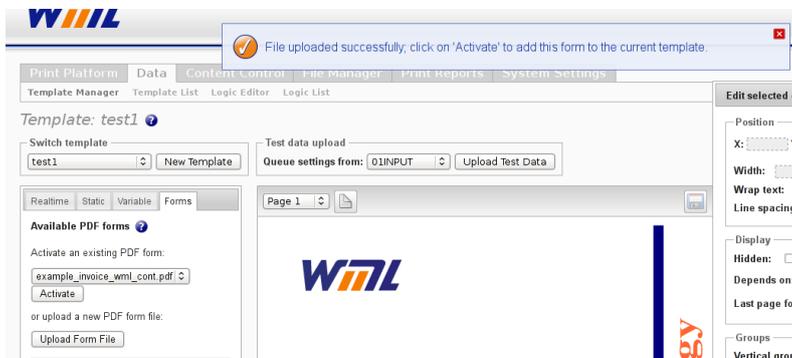


If your DataGateway already contains at least one template, you can create a new template from within the template manager, by clicking the **New Template** button (Figure 7.29).

Upload and activate PDF forms

Now that you have a template, it is time to upload a form. Select **Data > Template Manager** and choose the **Forms** tab in the overlay elements box (see Figure 7.21).

Figure 7.30. The PDF was uploaded and is now available



Press the button **Upload Form File** and use your browser's dialog box to select your PDF file. The file will be uploaded to the DataGateway. The DataGateway will parse the file to determine if it can successfully overlay the document onto another.

If the PDF form is correctly parsed by the DataGateway, a success message will read “File uploaded successfully; click on 'Activate' to add this form to the current template” (Figure 7.30).

Note

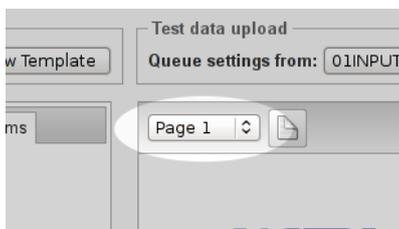
An uploaded PDF form may be added any number of times to any template on the DataGateway.

Figure 7.31. The PDF was uploaded but contained an unsupported feature



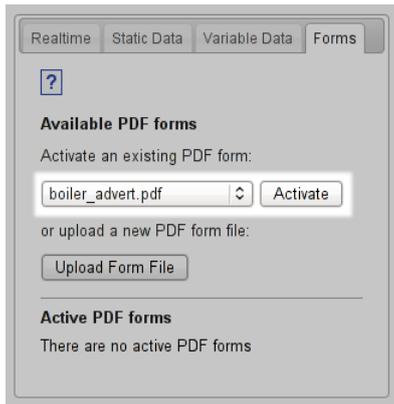
If the DataGateway encounters a known problem with your PDF form, it will respond with an error message describing the problem (Figure 7.31). In this case, please contact support with details of the problem.

Figure 7.32. Checking the template page



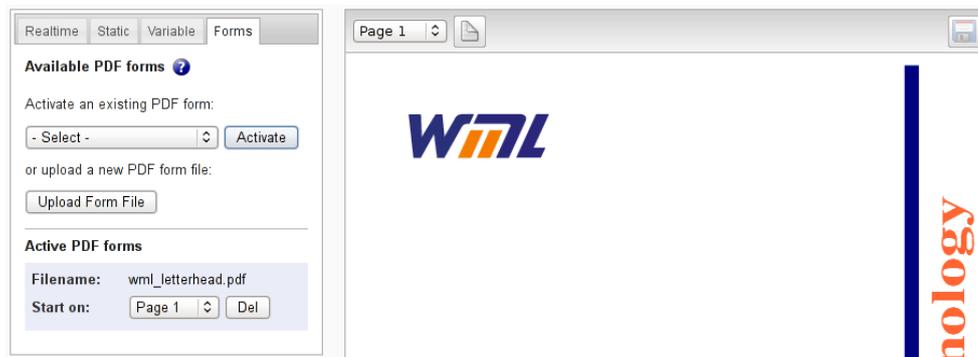
To activate your form, first ensure you are on the page on which the PDF form should be applied. See Figure 7.32 which shows the pages menu. When you activate a form, its first page will appear on the current template page.

Figure 7.33. Activating a form



Select the form from the menu highlighted in Figure 7.33. Press the **Activate** button. The form should appear on the canvas, as shown in Figure 7.34.

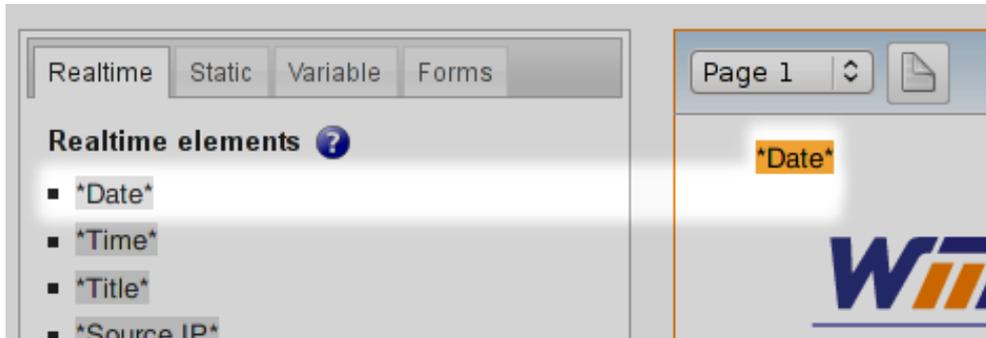
Figure 7.34. An active form showing on the canvas



Add text overlays

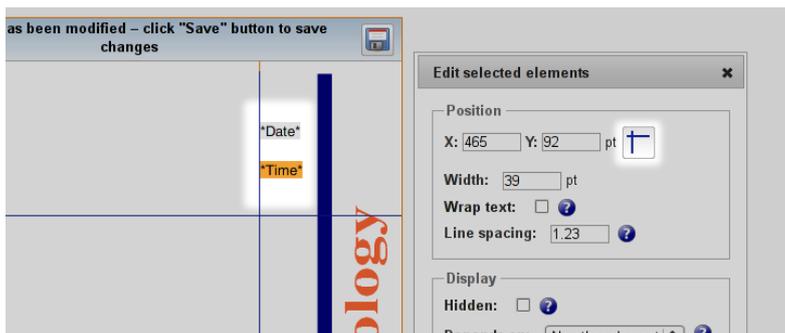
Now that a form has been added to the template, we'll add a date and time stamp to the example.

Figure 7.35. A date element is added to the template



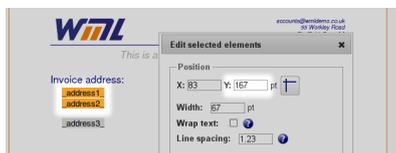
Dates and times are *realtime data* text elements. Select the **Realtime** tab. Drag a date element onto the canvas and position it as shown in Figure 7.35. Do the same for a time text element.

Figure 7.36. Using the snap-to guides to align the text elements



The template manager has a snap-to guide to help you align your text elements. Click the snap-to button (⊞) to turn on the guides. Drag the elements towards the guidelines and they will snap into position (Figure 7.36).

Figure 7.37. Moving multiple elements



You can select multiple elements and move them together. Click on the first element to select it. Now press the shift key and click on the second element. Both are now selected as shown in Figure 7.37.

To move the selected elements you can edit the text boxes in the element editor; in Figure 7.37 the Y position is being modified. Alternatively, use the cursor keys to “jog” the position of the selected elements.

When you are satisfied with the position of the text elements, press the save button to write the template to the DataGateway.

Review template settings

Figure 7.38. The template settings box

Template settings

Name

Basis

Num pages page

Copies extra copies

Elements Duplicate on copies

Page size

Format

Units

Date format

Time format

Before we test a print job, refer to the **Template settings** box (Figure 7.38). Check that the **Page size** is set to the same size as your input print data¹. If the page size here, and the page size of your input print data differ, then text elements may appear out of position when printed.

Finally, choose the format in which you would like the date to be printed, and the format for the time.

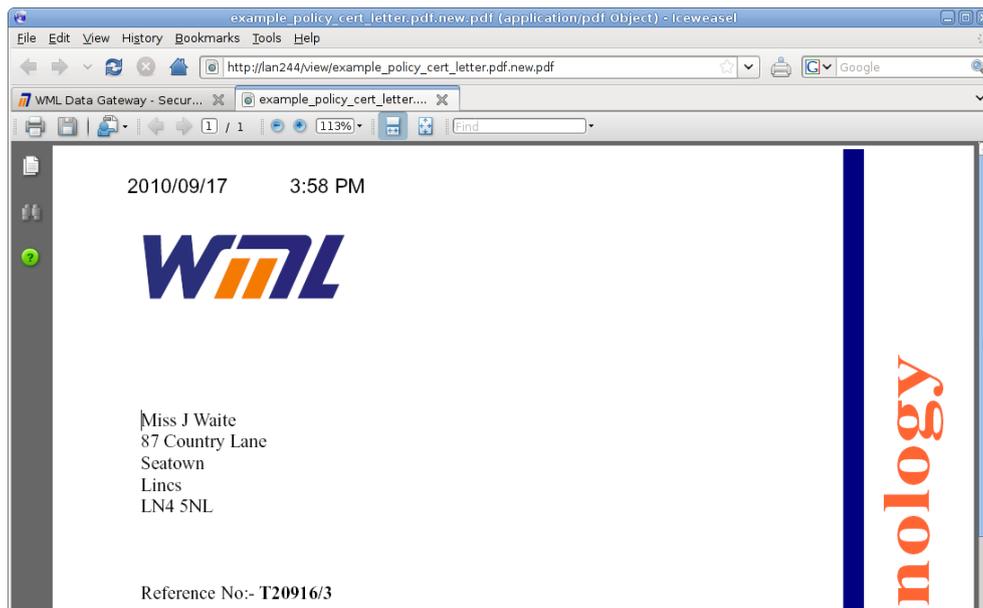
Upload and test

To find out how your input print data will be processed by the new template, use the test upload feature in the **Test data upload** box.

¹ Where applicable; if your input print data is structured text data such as CSV or XML formatted data, then choose the page size in which you wish the document to be created.

Select your new queue (here we called it “newQueue”) from the menu below **Queue settings from:**. Press the **Upload Test Data** button and use your browser's file dialog to select your input print data.

Figure 7.39. The test print job displaying in a new browser tab



The test print data will be uploaded onto the DataGateway, it will be processed through the current template, using any queue-specific settings from “newQueue” which will generate an output PDF file. Your browser will be re-directed to open the PDF file in a new tab or window (Figure 7.39).

Note

It's important to allow popups in your web browser. If you upload a test file, and no output PDF document appears this may be the cause.

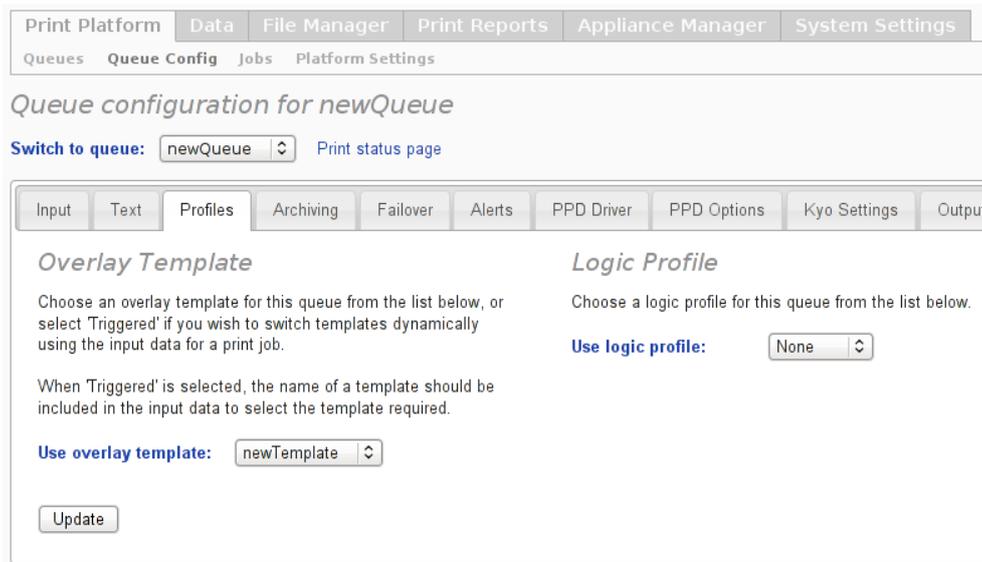
The example shown in Figure 7.39 shows the input print data (a letter to a Miss J Waite) merged with the PDF form (a WML letterhead) and the date and time realtime data elements, which have been printed as the date and time when the test data example was processed.

Bind the template to the queue

When you are satisfied with the PDF documents generated with your template, you are ready to print a document through your queue. Before printing, you must “bind” your template to the queue.

Enter the **Print Platform > Queue Config** page and switch to your new queue.

Figure 7.40. Binding a template to the queue



Select the **Templates** tab (Figure 7.40). The menu labelled with **Use overlay template** selects the template which will be applied to print jobs passing through this queue. Select your new template² and press the **Update** button.

Send your print data to the DataGateway using your preferred method (LPD, IPP, raw, hot-folder or HTTP POST) and verify that the document prints as expected.

7.7. Using structured input data

Structured input data is print data sent as a series of specific data items, the value of which may vary between print jobs, but which are to be printed in a pre-defined manner. For example, a database of sales records can be used to produce a set of data relating to each sale. Each data set can then be printed in the form of an invoice, by specifying the position of each data item on a blank invoice template.

The format for structured input data can be XML, comma-separated values (CSV) or Jet-forms compatible data. XML data is structured as explicit key-value pairs (and may contain `Job` and `Page` sections), while CSV data may be either a simple ordered list (separated by line breaks, commas or another chosen separator character), in which case the position in the list of an item is the means by which that item is specified, or key-value pairs separated by line breaks.

² For greater flexibility you may wish to trigger your template from words in your input print data. See Chapter 8 for more information.

Note

The structured input data code has been designed to make it straightforward to develop additional, customer-specific input data formats.

The type of structured data is detected automatically by the DataGateway as the print data is received and the data is laid out according to the rules defined in the overlay template assigned to the local queue or corresponding to a trigger value contained in the input data.

XML data is preferred over CSV, as although it appears a little more complex than simple CSV data, the layout is more flexible and errors in the input data are more likely to lead to *no* output rather than *incorrect* output. It is easier to detect that there is an error in the system if no document appears!

XML input data

In XML input data, the first line should be an *XML declaration*, the presence of which tells the DataGateway that the data is in XML format. The DataGateway looks for sections and key-value pairs within a parent XML element named `WmlXmlData`.

A single XML input data file may contain the variable data for either a single document, or a batch of many. The input data may be given an overall batch reference, with a `variable` element with `key` attribute equal to “`WmlBatchRef`” (the value of `WmlBatchRef` is encoded in the Info dictionary of the generated PDF).

Note

Within the context of structured input data, the term “job” is equivalent to “document” or “document data-set”.

Alongside the batch reference, it is also possible to specify the default “`WmlTemplate`” for all documents in the batch. These two variables are shown in Example 7.2.

Example 7.2. First lines of some XML input data

```
<?xml version="1.0"?>
<WmlXmlData>

  <!-- Batch reference string -->
  <variable key="WmlBatchRef" value="MyInputBatch" />
  <!-- Batch-default template -->
  <variable key="WmlTemplate" value="myTemplate" />
```

Job and Page elements

Within the parent `WmlXmlData` element, each document data-set, or “job”, should be enclosed within a `Job` element. Each job should be given an `id` attribute, which is the job number, and should start at 1 (rather than 0).

Warning

If the input data contains more than one job, then it is necessary to specify an output queue to process the output documents built from the input batch. This is set in **Print Platform > Queue Config > Text**.

Inside the job element, key-value pairs can be specified either to apply to the entire document, or to each page of the document. Key-value pairs are stored in `variable` elements, each of which defines a single piece of text to be placed on the print job. Each `variable` element should have a `key` attribute and a `value` attribute. The value will be placed and formatted on the document according to the location, font-face and font-size which has been assigned to the key in the Template Manager.

The special key-value “WmlTemplate” can be specified to override any batch-default template for the job. The key-value “WmlDocRef” is used to provide a unique document identifier, which is encoded into the Info dictionary of the generated PDF. “WmlDocRef” is used as the main document identifier within the Document Content Control module (see Chapter 9).

If your document needs to display a different value for a given key-value pair on each page, then you should separate the job with `Page` elements. Each `Page` element must contain an `id` attribute, specifying the page number. If you use `Page` elements, you can use the element with `id` equal to 0 for variables which apply across the job, as shown in Example 7.3.

Example 7.3. Page elements within a job with Page ID 0

```
<Job id="1">
  <Page id="0"> <!-- Page ID 0 means "this isn't a page,
                it's for the whole job -->
    <variable key="WmlDocRef" value="Report10002"/>
  </Page>
  <Page id="1">
    <variable key="invoice_num" value="125"/>
  </Page>
  <Page id="2">
    <variable key="invoice_num" value="126"/>
  </Page>
</Job>
```

Example 7.4 shows the document wide key-value directly within the `Job` element. Example 7.4 is exactly equivalent to Example 7.3. In both cases, this information is to be used to build a report on which the value for “invoice_num” is “125” on page 1 and “126” on page 2. The unique document reference for the job is “Report10002”.

Example 7.4. Page elements within a job

```
<Job id="1">
  <variable key="WmlDocRef" value="Report10002"/>
  <Page id="1">
    <variable key="invoice_num" value="125"/>
  </Page>
  <Page id="2">
    <variable key="invoice_num" value="126"/>
  </Page>
</Job>
```

If the template is specified with “WmlTemplate” for the entire document, then the template is expected to have as many pages as there are Page elements in the XML, or the template should have an **Others** page.

Note

See Section 7.5 for more detail about the Template Manager. The **Others** page in a WML Template specifies the layout for all pages which are not uniquely laid out. For example, you may specify a first page layout for a letter, a second page layout for some terms of business, then an **Others** page for all continuation pages after page 2.

You can use templates as “page templates”, rather than as document-wide templates. In this mode, you specify a WmlTemplate for each page in the job. Only the first page of each template is applied to the page of the document.

Example 7.5 shows a full, annotated example of an XML input batch file for multiple jobs.

Example 7.5. Full XML example

```
<?xml version="1.0"?>
<WmlXmlData>

  <!-- Batch reference string -->
  <variable key="WmlBatchRef" value="Input Batch Name"/>
  <!-- Batch-default template -->
  <variable key="WmlTemplate" value="someTemplate"/>

  <Job id="1">
    <Page id="0"> <!-- Page ID 0 means "this isn't a page,
                    it's for the whole job -->
      <variable key="WmlDocRef" value="Doc 1"/>
    </Page>
    <Page id="1">
      <variable key="key1" value="something1"/>
    </Page>
    <Page id="2">
      <variable key="key1" value="something1.1"/>
    </Page>
  </Job>

  <Job id="2"> <!-- NB: No page elements in this one,
                    page is implicitly equal to 0. -->
    <variable key="WmlDocRef" value="Doc 2"/>
    <variable key="WmlTemplate" value="overrideTemplate"/>
    <variable key="key1" value="something2"/>
  </Job>

  <Job id="3"> <!-- This job shows the use of per-page templates -->
    <variable key="WmlDocRef" value="Doc 3"/>
    <Page id="1">
      <variable key="WmlTemplate" value="CoveringLetter"/>
      <variable key="name" value="Bob"/>
    </Page>
    <Page id="2">
      <variable key="WmlTemplate" value="TermsOfBusiness"/>
      <variable key="date" value="2011-01-03"/>
    </Page>
  </Job>

</WmlXmlData>
```

Previous XML data format

An example of the older format for XML structured input data is shown in Example 7.6. This format lacks the flexibility to specify job and page elements, but the DataGateway will still accept this data.

Example 7.6. XML input data - old format

```
<?xml version="1.0" ?>
<!DOCTYPE WmlXmlData >
<WmlXmlData>
  <variable key="WmlTemplate" value="sales_contract"/>
  <variable key="WmlDocRef" value="REF123456"/>
  <variable key="copyLabel" value="Customer Copy"/>
  <variable key="refNo" value="7101-E9418"/>
  <variable key="company" value="CFG SUPPLIES"/>
  <variable key="address1" value="1 MINNS WAY"/>
  <variable key="address2" value="FAIRFORD RETAIL PARK"/>
  <variable key="address3" value="WESTFIELD, YORKS W1 9GT"/>
  <variable key="telephone" value="Tel: 01234 567890"/>
</WmlXmlData>
```

Character encoding and special characters

The DataGateway supports the following character encodings for XML input data:

- UTF-8
- UTF-16 (both little and big endian)
- ISO-Latin-1 (ISO-8859-1)
- ASCII

The default encoding is UTF-8. To use one of the other supported encodings, specify the encoding in the XML declaration as shown below.

```
<?xml version="1.0" encoding="iso-latin-1" ?>
```

A number of characters have special meaning in XML data. If you wish to include any of these characters in the print data you must use the corresponding character sequence shown in the following table.

Table 7.1. Special characters in XML

<	<
>	>
'	'
"	"
&	&

Note

The characters used for data keys should be confined to the ASCII character set and should not include any of the special characters above.

CSV data

Comma-separated format is simpler than XML, but a little more restrictive and also prone to error. The main drawback to using comma-separated values is that the character used as separator cannot be used within the values.

This restriction aside, separating values using commas, spaces or new-line characters is often adequate, and many applications generate output data in this format.

Note

To ensure that the DataGateway consumes CSV data as structured data, rather than plain text, set **Treat input text** to “as structured data” in **Print Platform > Queue Config > Text**.

DataGateway accepts data in two forms of CSV; “ordered” and “key-value”. Ordered data is simply a list of values, separated by a special character, with no keys in the listing (as described below, the keys are implied). Key-value data is a table of key-value pairs.

Note

The field separator character can be specified in **Print Platform > Queue Config > Text**.

DataGateway will auto-detect whether the data has been arranged as ordered CSV or in key-value pairs. Care must be taken to ensure that the separator character does not appear in either the keys or the values, otherwise unexpected and confusing results will be seen.

Ordered CSV

Ordered CSV is a simple list of values, one value following the next, separated either by new-lines or by the separator character. The new-line can be either the line-feed (0xa, LF or \n) character used in Unix (including Linux and Mac) systems or the DOS carriage-return/line-feed pair (0xd 0xa, CRLF or \r\n).

The ordered list has implied keys. Internally, DataGateway assigns this list the keys 0, 1, 2, 3, and so on. The ordered list can be separated by new-lines, with a new value on each new line, as shown in Example 7.7 or by a user-specified character such as a comma (default) or semi-colon (see Example 7.8). In both examples, the implied key “0” is “sales_contract”, key “1” is “Customer Copy” and so on.

Example 7.7. Column list CSV data

```
sales_contract
Customer Copy
CFG SUPPLIES
1 MINNS WAY
FAIRFORD RETAIL PARK
WESTFIELD, YORKS W1 9GT
Tel: 01234 567890
7101-E9418
```

Example 7.8. Row list CSV data

```
sales_contract;Customer Copy;CFG SUPPLIES;1 MINNS WAY;FAIRFORD RETAIL PARK;  
WESTFIELD, YORKS WS1 9GT;Tel: 01234 567890;7101-E9418
```

Key-value CSV

The format of key-value CSV data is shown in Example 7.9. The data is laid out as a table. The columns of the table are delimited by the user-specified separator character and the rows of the table by new-lines. The first, left-hand column contains the keys, the second column contains the values.

In the example, the key “address1” has the value “1 MINNS WAY” and key “telephone” has the value “Tel: 01234 567890”.

Note

Unfortunately, it is *not* possible to arrange the key-value table as two rows instead of two columns.

Example 7.9. CSV key-value data

```
WmlTemplate,sales_contract  
copyLabel,Customer Copy  
company,CFG SUPPLIES  
address1,1 MINNS WAY  
address2,FAIRFORD RETAIL PARK  
address3,WESTFIELD, YORKS W1 9GT  
telephone,Tel: 01234 567890  
refNo,7101-E9418
```

Specifying multiple jobs

To specify multiple jobs in a single CSV file, a string can be specified as a job separator. Any lines of the CSV input containing this string will cause one job to end and the next to begin. The job separator is specified in **Print Platform > Queue Config > Text**. If the input data contains more than one job, then it is necessary to specify an output queue to process the output documents built from the input batch. This is also set in **Print Platform > Queue Config > Text**.

Character encoding and special characters

Data should be encoded as ASCII or ISO-Latin-1 (ISO-8859-1). For key-value data, the key should be confined to the ASCII character set and should not include any of the characters <, >, ', ", or &.

Jetforms format input data

The DataGateway will accept incoming data in Jetforms format. For more information, please contact WML.

Template special keys

For all data formats there is a key (or position) that is reserved for specifying the template to be used for the current print job. In the following example, the queue has been set up to select the template by the “Triggered” option (the section called “Overlay templates and logic profiles”). If the value specified by the key corresponds to the name of an existing template then that is what will be applied to the current print job. In this way, structured print data passed through a single queue can be printed in a variety of ways according to which template is selected.

For key-value XML or CSV input data, the reserved key used to specify the template is “WmlTemplate”. If your input data is ordered CSV, then the first entry (implied key “0”) is used to specify the template. The example listings above all show the use of the reserved trigger key to specify the overlay set “sales_contract”

Also reserved are the keys “WmlDocRef” and “WmlSortKey”. If either is present, then its value will be inserted into the info dictionary of the PDF file as a metadata item. It is intended that the customer or document reference number or identifier would be placed into “WmlDocRef”.

The Document Content Control module makes use of the WmlDocRef for automatic document collation. WmlSortKey is used for any information which should be used for pre-output document sorting. For example, this key could hold the destination postcode for the document, and that information would then be used to mailsort the documents prior to printing.

Note

Unlike the WmlTemplate key, the WmlSortKey and WmlDocRef keys can *only* be specified in key-value data (XML or key-value CSV).

7.8. Using data groups

This section provides examples of using the data groups feature of the Template Manager.

Letter with variable content

Consider creating a template for a letter, the contents of which are not known in advance. In particular, the main body of the letter may comprise several paragraphs, the number and contents of which will affect the length of the resulting document.

We can use a *data group* element to indicate an area of the canvas to be filled with the paragraphs in the main body of the letter. A set of variable data elements will be specified for the group, and as many of these as occur in the input data will be added to the data group area, and laid out according to the data group's parameters.

Figure 7.41. Example showing a data group for the body of a letter

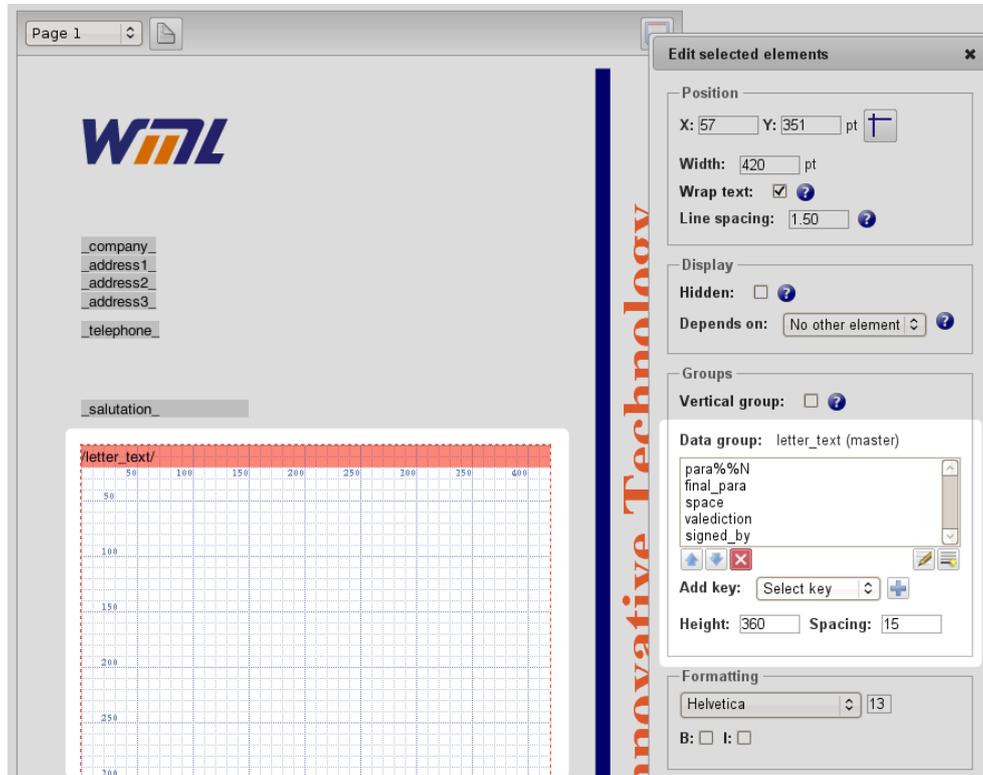


Figure 7.41 shows the canvas for our letter template, with a data group called “letter_text” to hold the main body of the letter placed beneath the “salutation” field.

The data groups controls in the element editor are highlighted in Figure 7.41. These fields are displayed only when a data group element has been selected. The name of the data group is given, followed by the word “master” in parentheses. This indicates that this is the master data group, for which we’ll specify the variable data elements to be displayed.

The multiple select box has been filled with several keys. These indicate the variable data to include in the group. Entries may be added to the list using the “Add key” dropdown below, which allows you to select from a list of example variable data keys or to specify the key text manually. The controls below the list can be used to delete or re-order the entries.

The first entry in the list is **para%%N**. The special tag %%N is used to indicate sequentially numbered elements. In this example, the entry indicates variable data elements with keys para1, para2 and so on. When an input job is processed, variable data elements with the corresponding keys will be added to the data group until the key for the next “N” is not found. At this point the next element in the list, if there is one, will be added.

The remainder of the keys displayed in this example are normal variable data keys. These are used to append the valediction and so on to the body of the letter.

Below the key list are fields for the height and spacing of the data group. The “Spacing” field governs the spacing between individual variable data elements in the group. Data group formatting also depends on width, text wrapping and line-spacing settings, which are found in the “Position” section of the element editor. Text wrapping is enabled by default for data groups.

Figure 7.42 shows the result of processing an input data file containing two variable data entries for the paragraphs in the main body of the letter. The two short paragraphs are added in sequence, followed by the closing lines of the letter, as specified in the variable data key list.

Figure 7.42. Output from the letter template

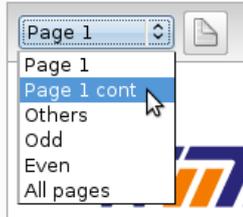


Now we need to handle letters for which the amount of text in the main body of the letter overflows the area specified for the master data group element, as in a longer letter with multiple paragraphs in the body text.

When a master data group is added to a template page, a new entry appears in the “Pages” dropdown menu (Figure 7.43). This entry is the *continuation page* for the data group, a tem-

plate page that can be set up to display data that overflows from the data group area on the normal template page. For page 1 of the template, the new entry will be “Page 1 cont”.

Figure 7.43. Pages menu showing continuation page entry



The continuation template page allows you to set up a page with a different layout to the original page on which to continue the display of elements in the data group.

Note

If no overlay elements are added to the continuation template page, a copy of the original page will be used as a continuation page instead.

To set up the continuation page to display the overflow from a data group, drag a copy of the data group from the “Existing data groups” list onto the continuation template page.

This creates a *continuation data group* for the original master data group. The variable data key list cannot be edited for this continuation group; the elements to be displayed are as specified for the master data group. However, the position, size and spacing for the continuation group may be changed as required.

You can add any other overlay elements, such as forms, static text or realtime data, to the continuation template page.

Figure 7.44. Letter template pages - original page (back) and continuation page (front)

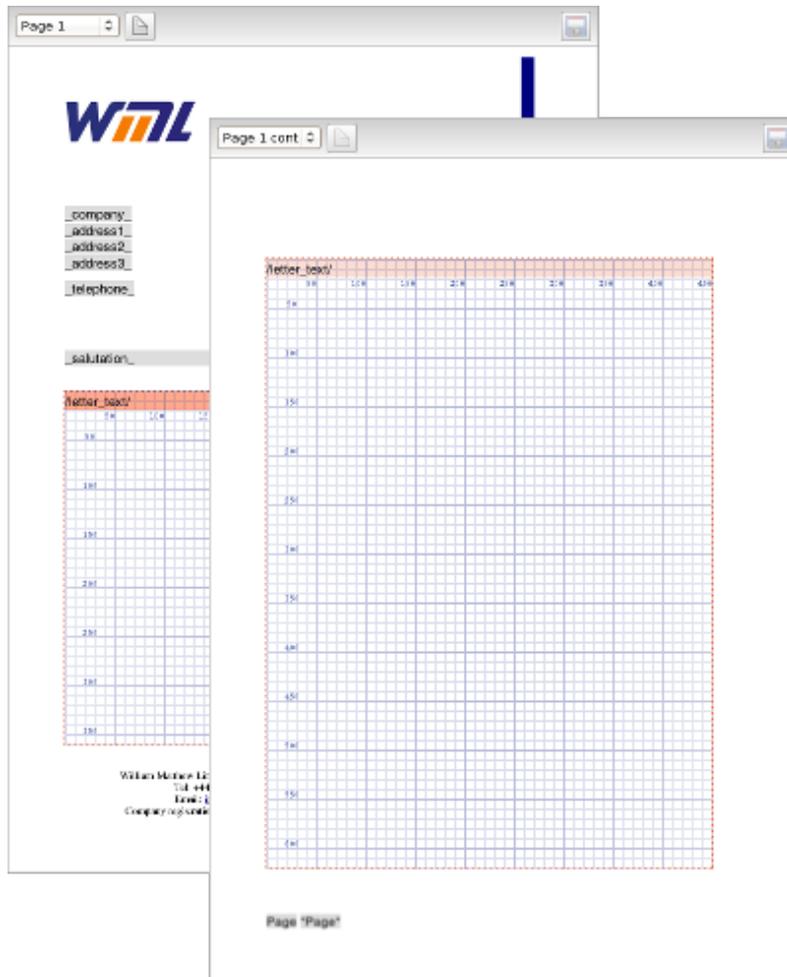
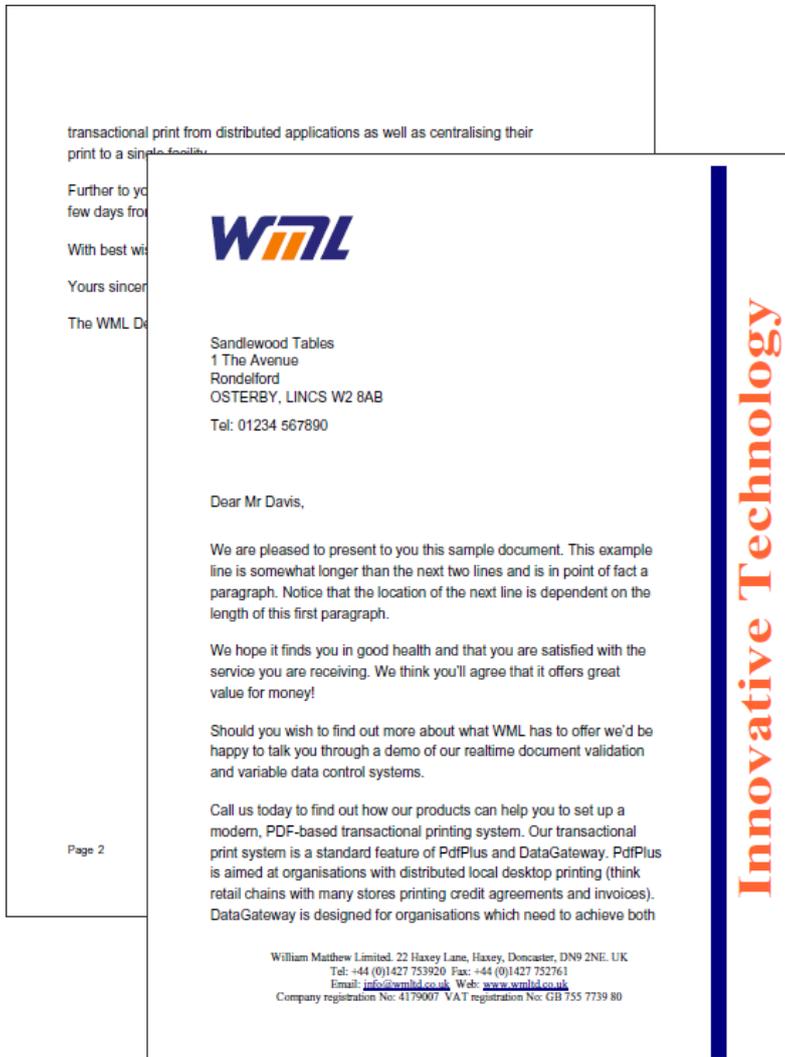


Figure 7.44 shows the continuation page for our example set up to display the remainder of the letter text on a plain page, starting at the top margin of the page and continuing down the full length of the page. Two text overlay elements have been added at the bottom of the page to display the current page number. Figure 7.45 shows the result of processing an input file for which the main text of the letter overflows onto the continuation page using this template.

In this case, the input data contains a long paragraph which overflows the space available in the master data group. The paragraph has been stopped where it reaches the bottom of the master data group. The remaining text in the paragraph has been placed on the continuation page, followed by the subsequent variable data elements. The page number text has been inserted with the correct page number.

Figure 7.45. Output from the letter template with overflow to continuation page



Multiple page invoice

In this example we look at using data groups to build a multiple page invoice. We introduce the concept of data group *columns* and show how these can be used to display table-based data.

Consider the invoice template shown in Figure 7.46. At the top of the invoice are standard variable data elements for the invoice address and reference numbers. Below is a table with columns for itemizing the invoice contents. Each item of the invoice will be represented

Template Management

by several variable data elements, which need to be displayed together as rows in the body of the table.

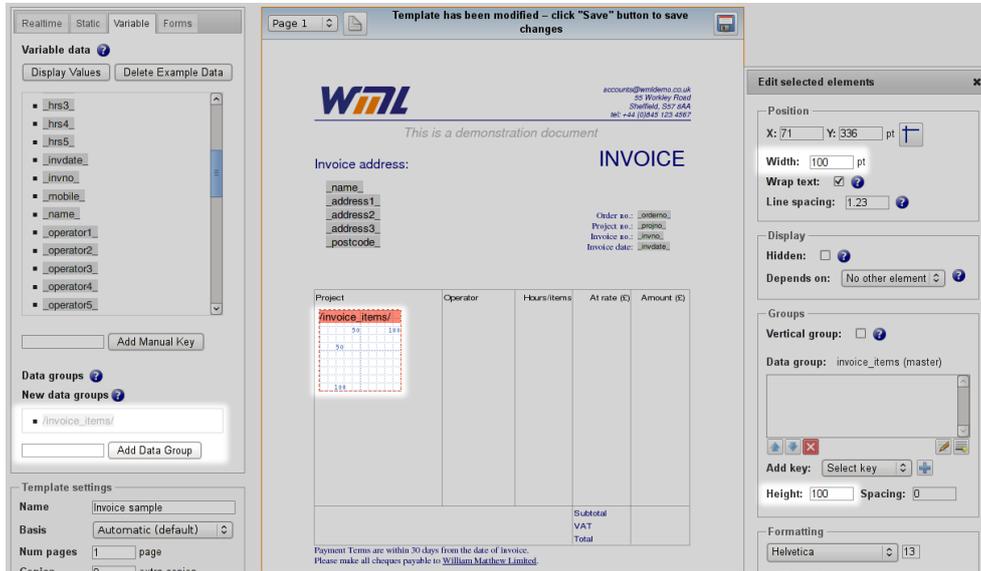
Figure 7.46. Example invoice template

The screenshot displays the Template Management interface for an invoice template. The interface is divided into several sections:

- Top Bar:** Includes 'Invoice sample' and 'New Template' buttons, and 'Queue settings from: 01INPUT' with an 'Upload Test Data' button.
- Left Sidebar:**
 - Variable data:** A list of variables including `_hrs3`, `_hrs4`, `_hrs5`, `_invdate`, `_invno`, `_mobile`, `_name`, `_operator1`, `_operator2`, `_operator3`, `_operator4`, and `_operator5`. Buttons for 'Display Values' and 'Delete Example Data' are present.
 - Data groups:** A section with an 'Add Data Group' button.
 - Template settings:** Fields for 'Name' (Invoice sample), 'Basis' (Automatic (default)), 'Num pages' (1 page), 'Copies' (0 extra copies), 'Elements' (checked 'Duplicate on copies'), 'Page size' (A4), and 'Format' (Default).
- Central Preview Area:** Shows a preview of the invoice template. It includes the WITL logo, contact information, and a table for invoice items. The table has columns for Project, Operator, Hours/Items, AI rate (£), and Amount (£). Below the table is a summary section with Subtotal, VAT, and Total. Payment terms are noted at the bottom: 'Payment Terms are within 30 days from the date of invoice. Please make all cheques payable to William Mathew Limited.'
- Right Sidebar (Edit selected elements):** Controls for the selected element, including:
 - Position:** X: 79, Y: 175 pt.
 - Width:** 46 pt.
 - Wrap text:**
 - Line spacing:** 1.23
 - Display:** Hidden: ; Depends on: No other element; Last page for: None.
 - Groups:** Vertical group:
 - Formatting:** Helvetica, size 13; Bold (B) and Italic (I) options.
 - Secure forms:** Mandatory: ; Auto-filled:
 - Remove** button.

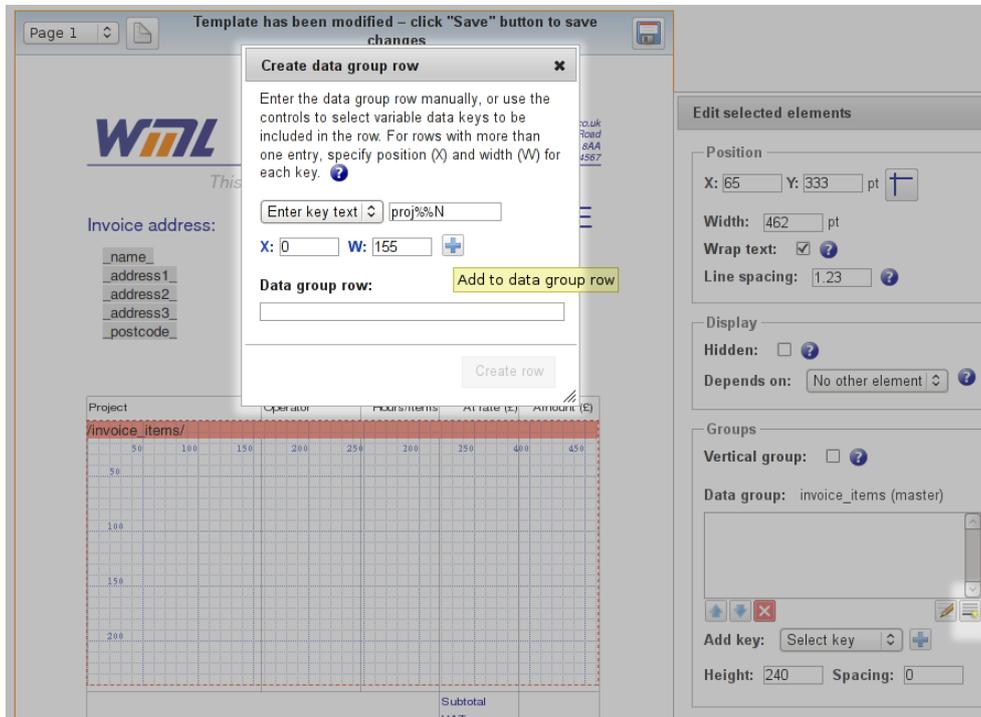
The first step is to add a data group for the invoice items (Figure 7.47). The newly added group has a default width and height, which we'll need to modify so that the data group corresponds to the area of the invoice table. This can be done using the width and height fields in the element editor, highlighted in Figure 7.47. Note that the data group area is displayed with a background showing gridlines. These will help us later with setting up the layout for the data group.

Figure 7.47. Adding a data group to the invoice template



For this example we need to create a row of variable data elements for each invoice item. We'll do this using the “Create data group row” controls, which are accessed using the  button at the bottom right of the data group key list (Figure 7.48).

Figure 7.48. Creating a data group row for the “invoice_items” data group



Each data group row may have a number of entries or *columns*. In the simplest case, the row is a single entry, and the only thing that needs to be specified is the variable data key that indicates the data value to insert. For tabular data, the row can be specified as a number of columns, each of which has a position and width within the row.

Figure 7.48 shows the controls for creating a data group row with columns. You can use the variable data key dropdown and the “X” (position) and “W” (width) fields to specify each column, using the add button (+) to add columns to the row. Alternatively, you can edit the data group row directly in the field below. The format for a data group row with columns is as follows:

key1,position1,width1,...,keyN,positionN,widthN

The keys may be specified explicitly or using the special tag `%%N` to indicate sequentially numbered elements (see the “Letter with variable content” example). The position is the horizontal distance in PDF units (points) of the column from the left hand edge of the data group. Similarly, the width specifies the width of the column in PDF units. You can use the gridlines displayed for the data group area to calculate appropriate values. When you have set up the required columns, click on “Create row” to add the row to the data group key list.

Figure 7.49 shows a document produced from an invoice template in which a data group row defines entries for the project, operator, hours, rate and amount columns of the invoice table. The entries are lined up in columns as specified in the data group row. Note that the project descriptions are wider than the column to which they belong, so they have been wrapped to fit.

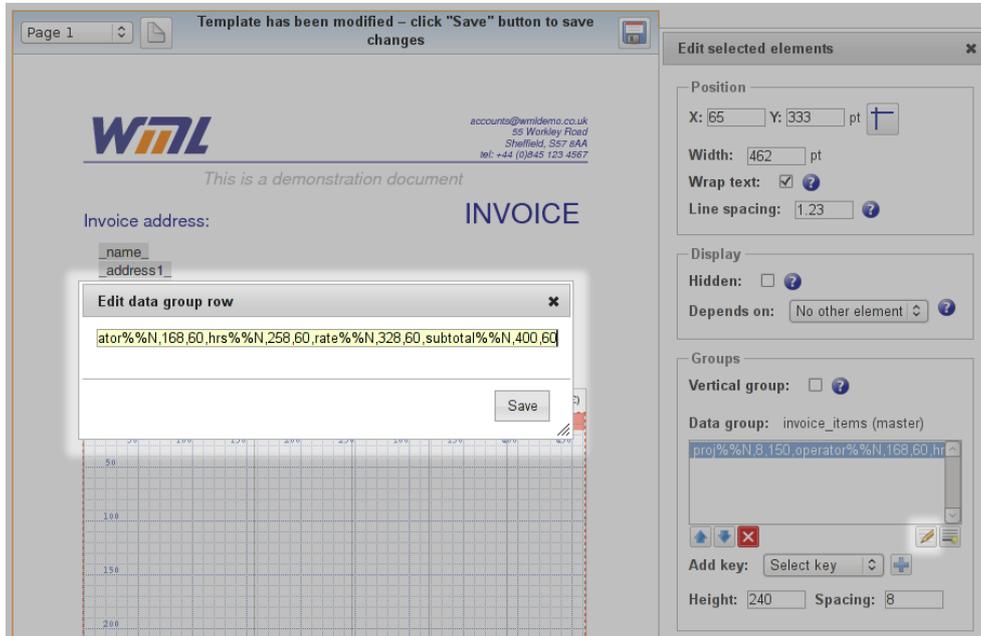
Figure 7.49. Sample output from the invoice template

The image shows a sample invoice document. At the top left is the WML logo. To the right is contact information: accounts@wmldemo.co.uk, 55 Workley Road, Sheffield, S57 8AA, tel: +44 (0)1945 123 4567. Below this is the text 'This is a demonstration document'. The word 'INVOICE' is prominently displayed in the upper right. On the left, under 'Invoice address:', is the address for FAO John Davies at 1 The Avenue, Rondelford, OSTERBY, LINCS, W2 8AB. To the right of the address are invoice details: Order no.: 243, Project no.: various, Invoice no.: 7438-9293, and Invoice date: 2013-02-11. The main body of the invoice is a table with five columns: Project, Operator, Hours/Items, At rate (£), and Amount (£). The table contains five rows of project data and a final row for the Subtotal.

Project	Operator	Hours/Items	At rate (£)	Amount (£)
Network infrastructure project A	A Smith	12	20.00	240.00
Network infrastructure project A	J Brendell	34	18.00	612.00
Network infrastructure project B	J Brendell	34	18.00	612.00
Network infrastructure project C	J Brendell	34	18.00	612.00
Network infrastructure project D	J Brendell	34	18.00	612.00
			Subtotal	

The positions of the columns in Figure 7.49 could be tidied up. To edit an existing data group row, highlight the row in the data group key list and click on the edit button (✎). This brings up a dialog where you can manually edit the data group row (Figure 7.50).

Figure 7.50. Editing a data group row



After modifying the row in accordance with the data group row format indicated above, click on “Save” to update the data group key list. Note that you'll also need to click on the save button (📄) to save the changes to the template.

Figure 7.51 shows the same document as before, with the column positions adjusted. The data group spacing, which governs the spacing between the rows, has also been increased.

Figure 7.51. Sample output with columns and spacing adjusted

accounts@wmldemo.co.uk
55 Workley Road
Sheffield, S57 8AA
tel: +44 (0)1945 123 4567

This is a demonstration document

INVOICE

Invoice address:
FAO John Davies
1 The Avenue
Rondelford
OSTERBY, LINCS
W2 8AB

Order no.: 243
Project no.: various
Invoice no.: 7438-9239
Invoice date: 2013-02-11

Project	Operator	Hours/items	At rate (£)	Amount (£)
Network infrastructure project A	A Smith	12	20.00	240.00
Network infrastructure project A	J Brendell	34	18.00	612.00
Network infrastructure project B	J Brendell	34	18.00	612.00
Network infrastructure project C	J Brendell	34	18.00	612.00
Network infrastructure project D	J Brendell	34	18.00	612.00
			Subtotal	

As in the “Letter with variable content” example, the document may need to be continued on one or more extra pages. Figure 7.52 shows a continuation template page for the invoice. This uses a background form with an extended invoice table. A continuation group for the “invoice_items” data group has been added to display any items which overflow from the master data group. The continuation group will use the same data group rows as specified for the master data group, so it will have the same column layout.

Additional variable data fields have been added to display subtotal, VAT and total figures for the invoice. Since these entries should come at the end of the invoice, they have been set as “Last page for” the “invoice_items” data group in the element editor, meaning that they will only be output when the last continuation page has been reached. (If there are no continuation pages, the elements will be output on the original page.)

Figure 7.52. Continuation template page for the invoice

The screenshot shows a software interface for managing invoice templates. The main window displays a template page titled "INVOICE - CONTINUED". At the top left is the WML logo. At the top right is contact information: accounts@wmldemo.co.uk, 55 Workley Road, Sheffield, S57 8AA, tel: +44 (0)1945 123 4507. Below this is the text "This is a demonstration document".

The central part of the page is a table with the following columns: Project, Operator, Hours/items, At rate (£), and Amount (£). The first row is labeled "invoice_items/" and contains values 5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, and 45.0. The table is mostly empty, with a "Subtotal" row at the bottom right. The "Subtotal" row has a yellow highlight. Below the table, there is a "Payment Terms" section: "Payment Terms are within 30 days from the date of invoice. Please make all cheques payable to William Matthew Limited." At the bottom of the page, it says "Registered Office: 55 Workley Road, Sheffield, S57 8AA. Registered in England No 1234 50 7VAT Reg. No 123 012345".

On the right side, there is a sidebar titled "Edit selected elements" with a close button (x). It contains several sections:

- Position:** X: 459, Y: 590, pt. Width: [] pt. Wrap text: [] [?]. Line spacing: 1.23 [?].
- Display:** Hidden: [] [?]. Depends on: No other element [?]. Last page for: invoice_items [?].
- Groups:** Vertical group: [] [?].
- Formatting:** Helvetica [?], 13. B: [] I: [].
- Secure forms:** Mandatory: [] Auto-filled: [].

 At the bottom of the sidebar is a "Remove" button.

Figure 7.53 shows the result of processing a data file for a longer invoice, which runs on to a second page.

Figure 7.53. Invoice with overflow to continuation page



accounts@wmldemo.co.uk
55 Workley Road
Sheffield, S57 8AA
tel: +44 (0)1645 123 4567

This is a demonstration document

Invoice address:

FAO John Davies
1 The Avenue
Rondelford
OSTERBY, LINCOS
W2 8AB

Project	Operator
Network infrastructure project A	A Smith
Network infrastructure project A	J Brendell
Network infrastructure project B	J Brendell
Network infrastructure project C	J Brendell
Network infrastructure project D	J Brendell
Network infrastructure project D	A Smith

Payment Terms are within 30 days from the date of invoice.
Please make all cheques payable to [William Mathew Limited](#).

This is a demonstration document

Registered Office: 55 Workley Road, Sheffield, S57 8AA.
Registered in England No 1234 567890



accounts@wmldemo.co.uk
55 Workley Road
Sheffield, S57 8AA
tel: +44 (0)1645 123 4567

This is a demonstration document

INVOICE - CONTINUED

Project	Operator	Hours/Items	At rate (£)	Amount (£)	
Network infrastructure project E	A Smith	5	20.00	100.00	
Network infrastructure project E	A Smith	5	20.00	100.00	
Network infrastructure project E	A Smith	5	20.00	100.00	
				Subtotal	3468.00
				VAT	606.90
				Total	4074.90

Payment Terms are within 30 days from the date of invoice.
Please make all cheques payable to [William Mathew Limited](#).

This is a demonstration document

Registered Office: 55 Workley Road, Sheffield, S57 8AA.
Registered in England No 1234 567890

Chapter 8. Logic Profiles

Logic profiles provide a way to trigger data manipulations on incoming print data *based on the content of the data*.

Each logic profile contains a number of *logic rules*. A logic rule contains a number of *logic conditions*. If these conditions are met by the incoming document, then the *logic action* is carried out.

Logic actions include:

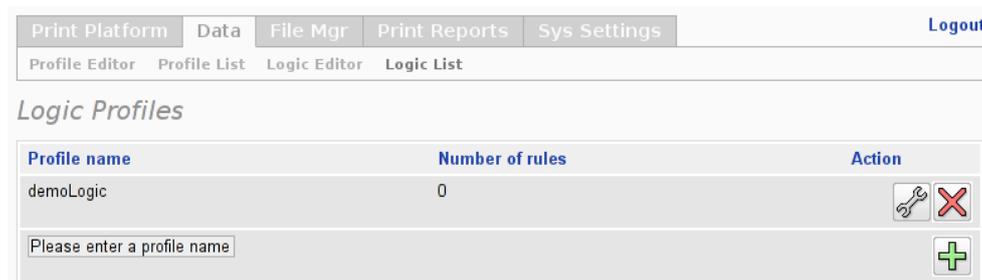
- Text search and replace
- Insert a barcode of the searched text at a specified location on the page.
- Trigger a template for the entire document.
- Trigger a PDF form for a page which meets the logic rule's conditions.

Figure 8.1. An empty logic profiles page



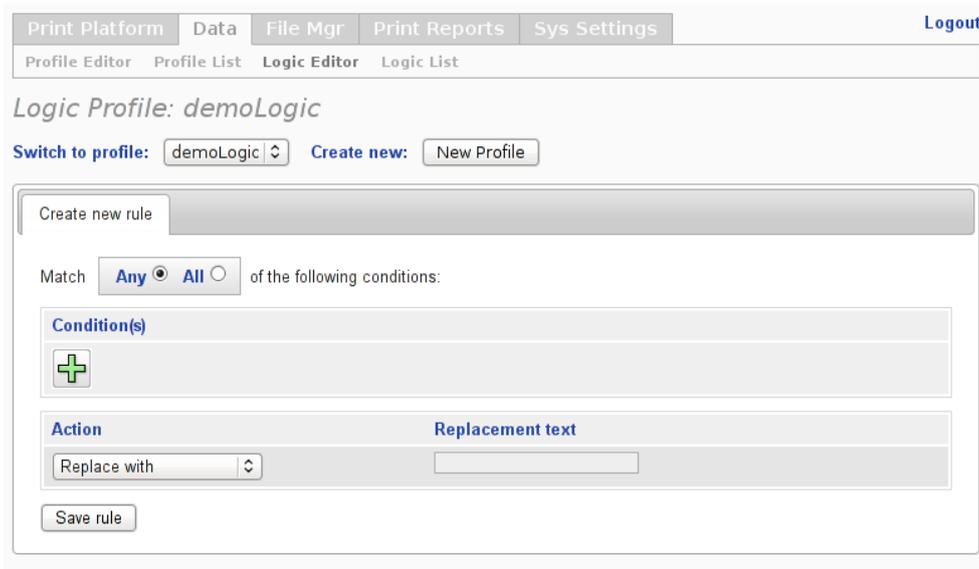
Logic profiles, like overlay templates, are managed using a logic profile editor and a logic profile list. To create the first logic profile, go to **Data > Logic List** (see Figure 8.1).

Figure 8.2. A new logic profile



Add a new profile by filling in the logic profile name, and pressing the add button (+). The new profile will appear in the list, as shown in Figure 8.2. Click on the configure button (🔧) to access the logic editor page for the new logic profile. The empty logic profile is shown in Figure 8.3.

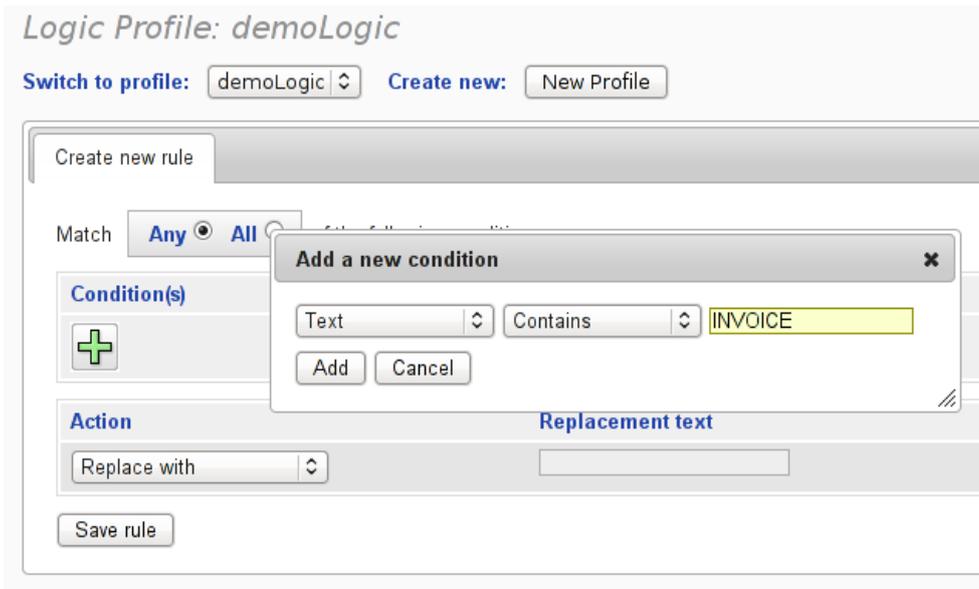
Figure 8.3. A new logic profile in the logic editor



In Figure 8.3, you can see a **Switch to profile** menu and a **New Profile** button, similar to those in the overlay template editor. These function in the same way.

The primary region of the logic editor is the rules box. In Figure 8.3 this shows a single tab headed **Create new rule**. Here you can add conditions to the rule and choose an action.

Figure 8.4. Adding a new condition to the rule



To create the new rule, you must add at least one condition and select the action.

To create a condition, press the add button (+). A dialog will pop up to create the condition. This is shown in Figure 8.4.

Note

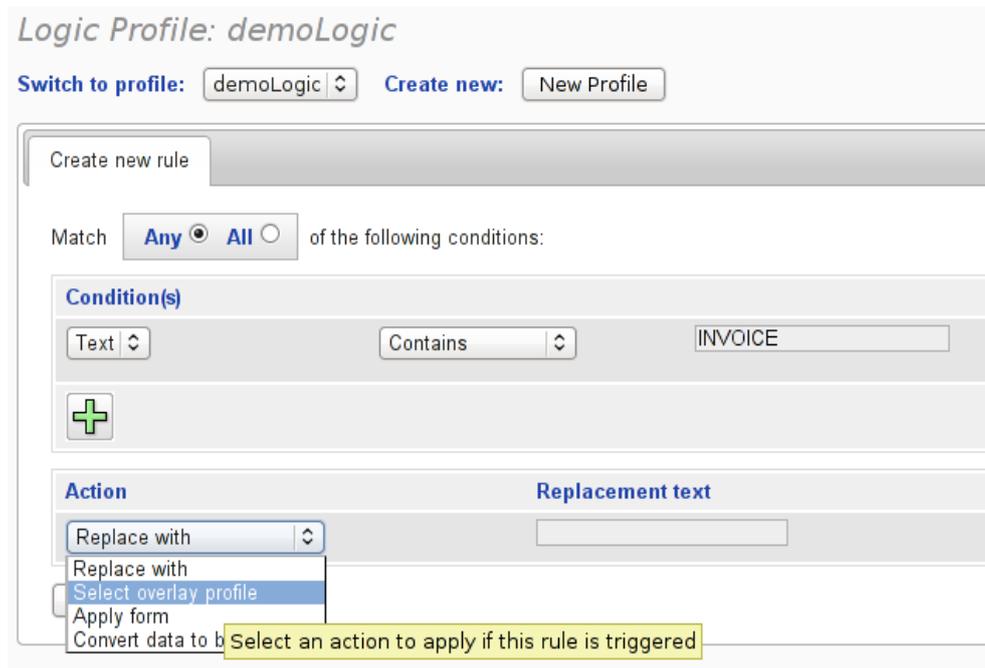
The left menu in Figure 8.4, whose value is “Text” currently has no other values as options. This menu is in place for future use.

Decide whether this condition should be “text contains” or “text does not contain”, and enter a match word or phrase. In the example shown, a condition that text should contain the word “INVOICE” is shown.

Note

All text conditions are case sensitive. If you need to match both upper and lower case, use two conditions and specify that **Any** condition must match.

Figure 8.5. Choose an action for the rule



Press the **Add** to add the condition to the rule. You can now add additional conditions and use the **Any/All** radio buttons to specify whether all of the conditions must match or whether it is sufficient for any condition to match.

Choose an action from the **Action** menu, as shown in Figure 8.5. The parameter for the action will depend on the action selected (Figure 8.6 and Figure 8.7).

Figure 8.6. Apply form - selecting a form

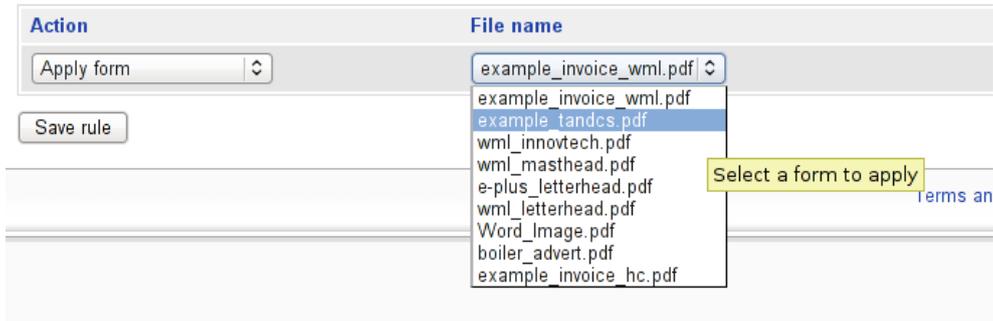
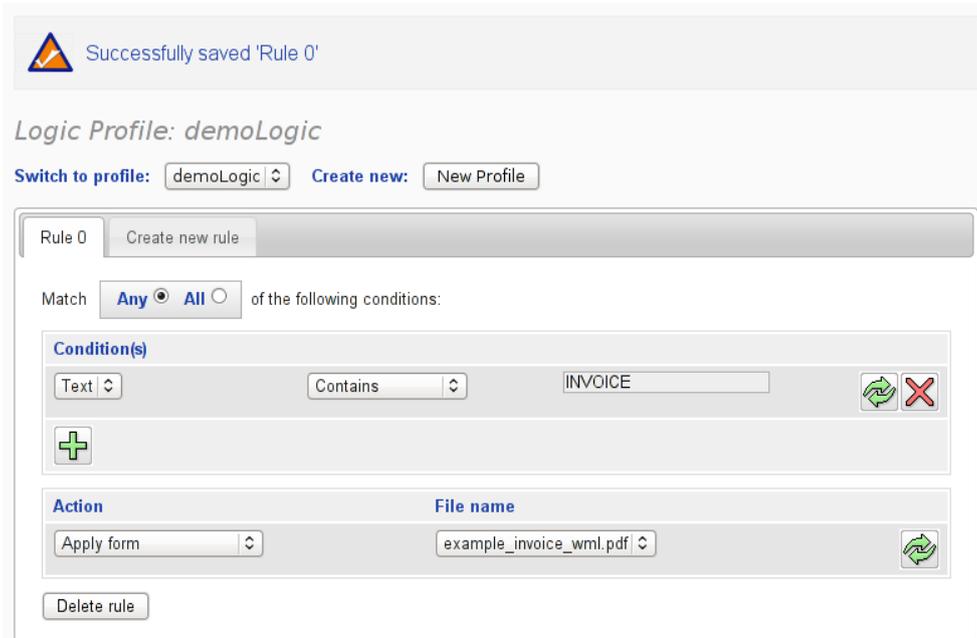


Figure 8.7. Select overlay template - available templates



Figure 8.8. The newly created logic rule



When you are satisfied, press the **Save rule** button. The tab will be re-named to **Rule0**, with a **Delete rule** button and a new **Create new rule** tab will appear. This is shown in Figure 8.8. The example rule will apply a PDF form called “example_invoice_wml.pdf” to any page of the input print data which contains the text “INVOICE”.

To make use of the logic profile, it must be *bound* to a queue in the same way that a template can be associated with a queue.

Figure 8.9. Binding a logic profile to a queue

The screenshot shows the 'Queue configuration for newQueue' page. At the top, there is a 'Switch to queue:' dropdown menu set to 'newQueue' and a 'Print status page' link. Below this is a navigation bar with tabs for 'Input', 'Text', 'Profiles', 'Archiving', 'Failover', 'Alerts', 'PPD Driver', 'PPD Options', 'Kyo Settings', 'Output', and 'Queue Info'. The 'Profiles' tab is active.

The main content area is divided into two sections: 'Templates and Logic' and 'Content Control'.
 In the 'Templates and Logic' section, there is an 'Overlay Template' subsection with instructions and a 'Use overlay template:' dropdown menu set to 'None'. Below that is a 'Logic Profile' subsection with instructions and a 'Use logic profile:' dropdown menu. This menu is open, showing 'None' and 'demoLogic' as options. An 'Update' button is located at the bottom of this section.

The 'Content Control' section has instructions and two dropdown menus: 'Use DCC profile:' set to 'New profile' and 'DCC output queue:' set to 'dccOutput'.

Access the queue configuration page at **Print Platform > Queue Config**. Select your profile from the the menu (Figure 8.9) and press **Update**. Print jobs on this queue (“newQueue”) will now be processed through the logic filter.

Chapter 9. Document Content Control

DataGateway provides a *document content control* interface to enable users to fine-tune the delivery of documents. Documents are analysed as they pass through the system, and any that are deemed, according to user-defined rules, to be incomplete or otherwise unfit for release are held back for review by system operators.

The document content control system extracts each document's unique identification details, so that sections of a document which have been delivered separately can be matched up. Other document content control features include ensuring that all required enclosures are present, checking for specified flags or exclusions, and performing validation checks, such as checking the form of a salutation or checking that a monetary value is within sensible bounds. The system can also extract information for use in other functions, such as electronic document delivery.

As an additional benefit, management information reports may be generated, to enable profiling and analysis of the numbers and types of documents passing through the system.

For more information about how DataGateway can meet your document content control requirements, please contact WML.

9.1. Overview

The document content control (DCC) system works by identifying the different *page types* that make up a document. Pages of certain user-specified types, called *document identifier pages*, are analysed to identify the overall *document type*.

Example 9.1. Document types

A document might contain the “letter” page type, amongst others. By analysing the contents of a letter page, it is possible to differentiate between a “quotation” document and an “order confirmation” document.

Once the document type has been determined, the document can be analysed further, according to its type, to determine its fitness for release.

A key feature of the system is the ability to ensure that a document is not sent to print until it is complete, that is, until every page that it should contain is present. This is essential when documents are split up across the incoming data and need to be combined as a single entity before they are released. The system uses a unique reference within the document to match up separate document portions. The document is collated when all of the required components are present.

Complete documents are subjected to a validation stage. A range of validation checks may be applied, such as checking dates, numeric or monetary values, the form of the salutation and address on a letter, and whether a value is present. Failure to meet the constraints of any validation check results in the document being held for inspection.

The system may modify incoming documents as necessary before they are released. For example, it is possible to add an out-for-mail barcode based on the document reference. The system can also be configured to collect data to be included in management information reports.

Data collected by the system is stored in the DCC database (**wmldcc**). This database is used to produce *DCC reports*, and may also be accessed and used to create customized management information reports.

9.2. Configuration

To configure the DCC system, select the **Content Control > Profiles** item from the main menu of the DataGateway web user interface.

In this section you can create and configure document content control *profiles*. Document content control is activated by applying a DCC profile to a print queue. You can maintain several profiles to be applied to different queues.

If this is the first time you have visited this section, you will be prompted to create a new profile. Otherwise, you should make sure that you have selected the profile that you wish to work on before continuing. You can change to a different profile using the **Switch profile** dropdown.

Profile configuration

Adding a profile

Click on **New Profile**, or select “Create new...” from the **Switch profile** dropdown. You will be prompted to enter a name for the profile. You should choose a name that differentiates the profile from any others on the system. You will also be asked if you wish to create a *core profile* (see “Core profiles and inheritance”).

Profile status

A profile must be *published* before it can be used in production. Any changes to the configuration of a profile will remain unpublished until you choose to publish the changes. This allows you to test modifications to the configuration before applying them to a production print queue.

Figure 9.1. Profile status



The **Profile status** box (Figure 9.1) displays the published status of the profile, and provides controls to publish, or revert, any unpublished changes. A new profile will display an initial status indicating that it has not yet been published.

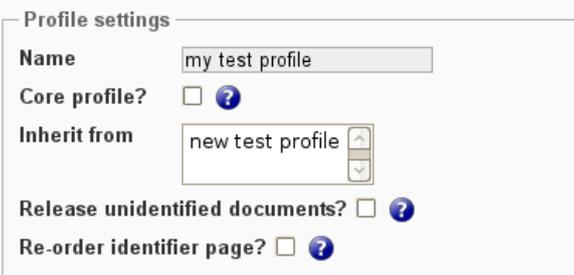
If you choose to publish the profile, your changes to the configuration will go live immediately. The published date will be updated to the current date and time.

If you choose to revert the profile, any changes since the last published date will be discarded.

Profile settings

You can modify the details of an existing profile by editing the fields in the **Profile settings** box (Figure 9.2).

Figure 9.2. Profile settings



The screenshot shows a 'Profile settings' form with the following fields:

- Name:** A text input field containing 'my test profile'.
- Core profile?:** A checkbox that is unchecked, followed by a blue question mark icon.
- Inherit from:** A dropdown menu showing 'new test profile' with up and down arrow icons.
- Release unidentified documents?:** A checkbox that is unchecked, followed by a blue question mark icon.
- Re-order identifier page?:** A checkbox that is unchecked, followed by a blue question mark icon.

Core profiles and inheritance

You can designate the current profile as a *core profile* by checking the **Core profile** setting. A core profile may be used to define a core set of properties, such as page types, that will be the same across a number of specialised implementations.

Other profiles can *inherit* properties from a core profile, meaning that it is not necessary to redefine common elements for each specialised implementation.

You can set a non-core profile to inherit from one or more core profiles, if any exist, with the result that the settings in the selected core profile(s) will take effect as if they were in the current profile.

Releasing unidentified documents

The **Release unidentified documents** setting allows you to relax the rules for holding unidentified documents.

If you choose to release unidentified documents, the system will allow documents to pass *whether or not a document type has been allocated*, as long as a document identifier page is present.

This means that documents can pass through the system even if a positive identification is not obtained. This may be useful if there are many varieties of document passing through the system and you do not wish to create a document type for each of them.

Note

Allowing documents to clear without identification reduces the document collation capabilities of the system, because it is necessary for a document type

to be assigned before it is possible to determine whether a given document is complete.

Re-ordering identifier pages

The **Re-order identifier page** setting determines whether or not to re-order the pages of a released document to bring a document identifier page to the front.

If you choose to re-order the identifier page, the system will bring the first identifier page to the front of the document, by moving the pages preceding it to the back of the document. This is useful if the identifier page is a page that should always be at the front of the finished document, such as a covering letter.

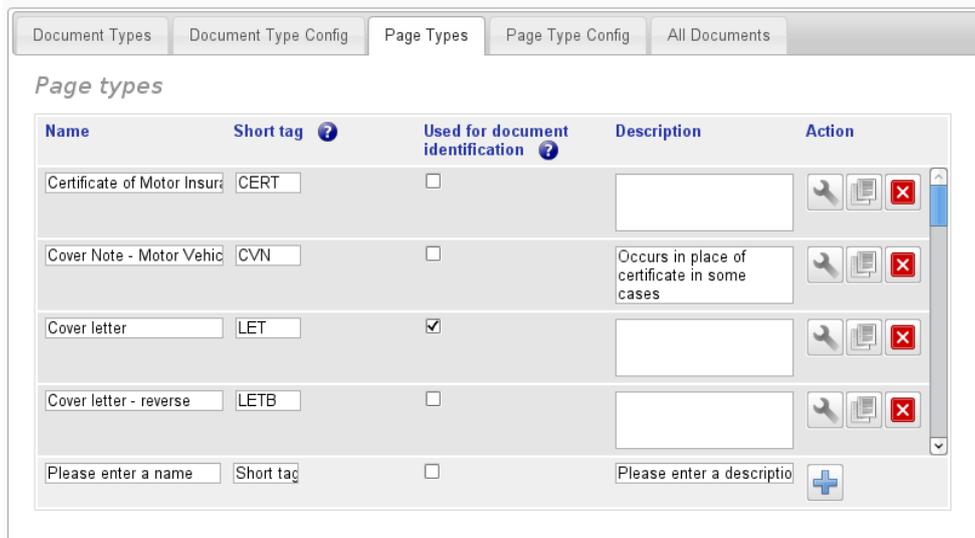
Page types

The first stage of document identification is to make a list of the *page types* occurring in your documents, and give the DCC system a set of rules that it can use to identify them.

Adding page types

Go to the **Page Types** tab to start adding page types to your profile (Figure 9.3).

Figure 9.3. Page types



You should define a page type for each distinct set of pages within your documents, for example, letter, terms and conditions, order details. Each page type is assigned a name, a short tag (used to abbreviate the page type in reports), and, optionally, a description. You can also specify whether a page type should be used for document identification (see “Document identifier pages”). After filling in the details for your new page type, click on the add button (+) to add it to the list.

The details of an existing page type can be modified at any time by editing its entry in the list of page types.

If you want to model a new page type on an existing page type, click on the duplicate button (📄) in the row for the existing type. This will create a copy of the chosen page type, which you can modify as appropriate for your new page type.

Note

When you duplicate a page type, DCC actions are not copied to the new page type. You will have to re-create DCC actions manually in the new page type.

Document identifier pages

At least one page type must be designated to be used for document identification. The contents of pages of this type, known as *document identifier pages*, will be analysed to differentiate between document types. If a document doesn't contain any identifier pages then it will not be possible to identify the document, so it is important to select an appropriate page type for this purpose. More than one page type can be designated to be used for identification; the combined contents of each such page in a document will be analysed during document identification.

Page type configuration

To configure a page type, click on the configure button (⚙️) for its entry in the list of page types, or go to the **Page Type Config** tab and select the page type from the **Switch page type** dropdown.

Figure 9.4. Page type configuration

The screenshot shows the 'Page Type Config' tab for a page type named 'Cover Letter'. The interface includes a 'Page type settings' section with fields for Name, Short tag, Used for document identification?, and Description. A 'Switch page type' section contains a dropdown menu and a 'New Page Type' button. Below this is a 'Page type identification phrases' table with columns for Phrase, Match case?, Strength, Notes, and Action.

Phrase	Match case?	Strength	Notes	Action
Dear	<input type="checkbox"/>	High		
Your Ref	<input type="checkbox"/>	Medium		
Yours sincerely	<input type="checkbox"/>	High		
Please enter an identification phrase	<input type="checkbox"/>	High		

Page type identification phrases

Page types are identified by means of user-specified *identification phrases* (Figure 9.4). These are phrases which may be linked, positively or negatively, with a given page type. By defining a set of identification phrases, and specifying the *strength of association* of each with the page type, it is possible to use the occurrence of these phrases on a page to determine whether it is of the given type.

Note

To keep your profile streamlined, try to use as few phrases as possible to identify each page type.

To enable a page type to be identified, populate the list of page type identification phrases with phrases that are likely (or unlikely) to appear on pages of this type. Try to select phrases that are specific to the current page type, to avoid getting “hits” for the phrases on pages of other types. You can choose to match the case of the phrase to make it more specific. Without case matching, you can expect more “hits” for some phrases. The notes field allows you to record information about why a phrase has been included, where it is expected to occur, and so on. This information will not have any effect on the analysis.

Strength of association

For each identification phrase, you must indicate the *strength of association* of the phrase with the current page type. It is useful to think of this in terms of whether the phrase is a strong, weak, neutral or negative identifier for the page type. A phrase that will only ever occur on this page type is a *strong* identifier; one that is likely to occur on other page types as well as on this type is a *weak* identifier. A phrase which is unlikely to occur on this page type is a *negative* identifier.

Note

When choosing identification phrases that are positive identifiers for a page type, it is useful to choose phrases which are unlikely to appear on other page types, to avoid “false positive” results for pages of other types.

Identification phrases that are negative identifiers for a page type can be used to disambiguate that type from other similar page types.

You can use the *neutral* setting to test the performance of the configuration without the contribution from a particular phrase. If you decide that an identification phrase is unnecessary, then it is best to remove it, as this will make the page type analysis more efficient.

By refining the strengths of the identification phrases associated with a page type, it is possible to fine-tune the identification of pages of that type within a document (see “Testing the configuration”).

Page type DCC actions

DCC actions may be set up to perform validation and other tasks on pages of specified types. DCC actions may also be applied to pages of all types, by configuring them on the **All pages** tab.

A DCC action implements a document check, a manipulation, or collection of positioned and labelled values.

The DCC actions available for page types are:

- Add PDF overlay
- Add PDF overlay on logic
- Add out-for-mail barcode
- Apply media controls
- Check address
- Check date range
- Check monetary value
- Check numeric value
- Check salutation and address
- Detect phrase flags
- Disallow phrases
- Ensure field is filled
- Get labelled document reference
- Get labelled value
- Get positioned value
- Get value array
- Hold on logic
- Set output queue
- Store PDF size

Please refer to “DCC actions” for information about individual DCC actions.

To add a DCC action, select the required action from the dropdown box. If the action requires parameters, a **Parameters** field will appear, together with text indicating the parameters that you should provide. Enter the parameters and click on the add button (+) to add the entry to the list of DCC actions.

Note

The initial parameters are mandatory; these must be entered in a specific order, and are specified as values only. Optional parameters follow; these may be entered in any order, and should be specified as key-value pairs (e.g. *matchcase=yes*). Parameters should be separated by commas.

An existing DCC action can be modified at any time by editing its entry in the list of actions.

Document types

The document content control system uses the contents of the *identifier pages* of a document to determine the *document type*. This stage of configuration requires you to make a list of the different types of document that you expect, and give the system a set of rules that it can use to identify each of them.

Adding document types

Go to the **Document Types** tab to start adding document types to your profile. You should define a document type for each distinct type within your documents, for example, quotation, order confirmation, delivery note. Each document type is assigned a name, and, optionally, a description. After filling in the details for your new document type, click on the add button (+) to add it to the list.

The details of an existing document type can be modified at any time by editing its entry in the list of document types.

Document type configuration

To configure a document type, click on the configure button (⚙️) for its entry in the list of document types, or go to the **Document Type Config** tab and select the document type from the **Switch document type** dropdown.

Document type identification phrases

Document types are identified by means of user-specified *identification phrases*. These are phrases which may be linked, positively or negatively, with a given document type. By defining a set of identification phrases, and specifying the *strength of association* of each for the document type, it is possible to use the occurrence of these phrases on an identifier page to determine whether the document is of the given type.

Note

To keep your profile streamlined, try to use as few phrases as possible to identify each document type. Note, however, that to ensure strength of identification, the system may require at least two “hits” for document type identification.

Enter phrases that are likely (or unlikely) to appear on the identifier page (or pages) of the current document type. Try to select phrases that are specific to the document type. You can choose to match the case of the phrase to make it more specific. Without case matching you can expect more “hits” for some phrases. The notes field allows you to record information about why a phrase has been included, where it is expected to occur, and so on. This information will not have any effect on the analysis.

Strength of association

For each identification phrase, you must indicate the *strength of association* of the phrase with the current document type. It is useful to think of this in terms of whether the phrase is a strong, weak, neutral or negative identifier for the document type. A phrase that will only ever occur on the identifier page for this document type is a *strong* identifier; one that is likely to occur on the identifier page for other document types as well is a *weak* identifier. A phrase which is unlikely to occur on the identifier page for this document type is a *negative* identifier.

Note

When choosing identification phrases that are positive identifiers for a document type, it is useful to choose phrases which are unlikely to appear on the identifier page for other document types, to avoid “false positive” results for other document types.

Identification phrases that are negative identifiers for a document type can be used to disambiguate that type from other similar document types.

You can use the *neutral* setting to test the performance of the configuration without the contribution from a particular phrase. If you decide that an identification phrase is not necessary then it is best to remove it, as this will make the document type analysis more efficient.

By refining the strengths of the identification phrases associated with a document type, it is possible to fine-tune the identification of documents passing through the system.

Associated pages

You can specify a list of *associated pages* (Figure 9.5) for a document type. These are page types that should be present in a document of this type when certain phrases occur on the identifier page. For example, a letter may state that the recipient should find enclosed the terms of business agreement; this statement can be used to indicate a requirement for pages of that particular type. If a page of a required type is not present in the document then the document will be marked as incomplete by the system, and will be held back to await completion or review.

Figure 9.5. Associated pages

Associated pages

Specify page types that should appear in a document of this type, based on phrases found on the document identifier page.



Associated pages are added by entering a phrase to trigger the association, and selecting one or more page types to be associated with the document when that phrase is found on the identifier page. Multiple page types can be selected from the list of available page types by holding down the control key while you make your selection.

If more than one page type is associated with a phrase, you can choose whether *all* or *any* of the selected page types should be present when the phrase is found.

Click on the add button (+) to add the entry to the list of associated pages.

Document type DCC actions

DCC actions may be set up to perform validation and other tasks on documents of specified types. DCC actions may also be applied to documents of all types, by configuring them on the **All documents** tab.

A DCC action implements a document check, a manipulation, or collection of positioned and labelled values.

The DCC actions available for document types are:

- Add PDF overlay on logic
- Apply media controls
- Associated pages on logic
- Disallow phrases
- Electronic document delivery
- Hold on logic
- Set output queue
- Store PDF size
- Tag output queue
- Trigger output queue

The DCC actions that may be applied to “all documents” are as above, with the addition of the following:

- Labels must match
- Labels must not match

Please refer to “DCC actions” for information about individual DCC actions.

To add a DCC action, select the required action from the dropdown box. If the action requires parameters, a **Parameters** field will appear, together with text indicating the parameters that you should provide. Enter the parameters and click on the add button (+) to add the entry to the list of DCC actions.

Note

The initial parameters are mandatory; these must be entered in a specific order, and are specified as values only. Optional parameters follow; these may be entered in any order, and should be specified as key-value pairs (e.g. *matchcase=yes*). Parameters should be separated by commas.

An existing DCC action can be modified at any time by editing its entry in the list of actions.

DCC actions

DCC actions provide a mechanism for performing a range of operations on completed documents, including validation, manipulation and information-gathering.

Some sample DCC actions are:

- *Add a PDF overlay*. This is a manipulation action. For example, a letterhead PDF is overlaid on pages of type “Covering Letter” to make a document ready for colour printing or electronic delivery
- *Check date range*. This is a validation action. This could be applied, for example, to ensure that the renewal date on a renewal invitation document is within a specified range.
- *Get labelled value*. This is an information gathering action. For example, the value “ABC corp” for the label “Underwriter” is collected from the document and stored for inclusion in a Management Information Report.

DCC actions have a plugin-style architecture, which means that customized actions can be developed quickly.

Applying DCC actions

There are a range of built-in DCC actions. Each is applicable to a page type, a document type, or both. Actions may be applied to the configuration for a given page or document type, or to the configuration for all pages or all documents.

DCC actions are controlled by providing a number of *parameters*. When you select an action that requires parameters, a **Parameters** field will appear, together with text indicating the parameters that you should provide. The initial parameters are mandatory; these must be entered in a specific order, and are specified as values only. Optional parameters follow; these may be entered in any order, and should be specified as key-value pairs (e.g. *matchcase=yes*). Parameters should be separated by commas.

Example 9.2. DCC action parameters

The **Get labelled value** action has one mandatory parameter, *Label*, and a number of optional parameters, which may be used to refine the position of the value with respect to the label and to control how the information is stored.

A parameter string of the form

```
Type of cover,addtodoc=yes,storeemptyvalue=yes
```

indicates that the value of interest is labelled with the text “Type of cover”, and that the value should be stored in the PDF information dictionary, even if it is an empty value. Since no position parameters are provided, the default values for these parameters will be used.

Logic-based DCC actions

Logic-based DCC actions have a *DCC logic expression* as their first parameter.

The logic expression is evaluated using information gathered from the document, such as labelled values, flagsets, page types, document types, or the processing queue name.

The operators *and*, *or*, *not*, *true*, *false*, *(*, and *)* are used to combine elements such as *mylabel=somevalue*, *dccpage3type=somepagetype*, and *dccdoctype=somedoctype*, to create the logic expression to be evaluated.

The components available for use in DCC logic expressions are:

`mylabel=somevalue` Test a labelled value. Returns true if the value with label `mylabel` is equal to `somevalue`.

`dccpageNtype=somepagetype` Test that page `N=1,2,3...` has the type `somepagetype`.

`dccdoctype=somedoctype` Test that the document has the document type `somedoctype`.

An example expression is:

```
dccdoctype=letter and insurance=car
```

If spaces are needed in an element, then use the escape character `\` before the space, or enclose the element in quotation mark characters `"`. For example:

```
dccdoctype=Covering\ Letter and "insurance=motor car"
```

Note

Enclose the *entire element* with enclosure characters: use `"label=element with spaces"` rather than `label="element with spaces"`.

To test for a missing value, leave the right hand side of the test expression empty. For example:

```
dccdoctype=Covering\ Letter and "dccdocref="
```

Normally, matches must be exact. If more complex matching is required, set `regex=yes` in the parameter list and use a regular expression. For example:

```
dccdoctype=^Covering.*$ and "insurance=^exactMatch$"
```

Other parameters to a logic-based DCC action follow the initial logic parameter.

Built-in DCC actions

The built-in DCC actions available on the system are described in the following sections.

A brief description of the function of each action is given, followed by a breakdown of the parameters to the action, if any. Mandatory parameters are listed first, in the order in which they should be applied, followed by optional parameters. Finally, the component to which the action may be applied is indicated.

Add PDF overlay

This DCC action uses the value associated with a label to select a PDF overlay for the page.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

If the labelled value is found, the system will select and apply the PDF overlay with name matching the value, appended by `".pdf"`, if it exists. For example, if the selected value is

“form1”, then “form1.pdf” will be selected from the PDF form files directory. If the file does not exist then the action will fail, resulting in the document being held for review.

Parameters

The parameters for the **Add PDF overlay** action are as detailed for the **Get labelled value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Add PDF overlay on logic

This DCC action uses logic to determine whether a specified PDF should be overlaid on the page.

A logic expression is evaluated, based on the value of detected labelled values, flagsets, page types, document types, or the processing queue name. If the logic evaluates to true on a given page, then a specified PDF is overlaid onto the page.

Parameters

The parameters for the **Add PDF overlay on logic** action are:

Logic expression	See “Logic-based DCC actions”.
<i>pdfoverlay</i>	(Optional) File name of PDF overlay to apply if logic evaluates to true.
<i>pdfoverlayback</i>	(Optional) PDF overlay to apply to the next page, if the document has an additional page.
<i>usepagevalues</i>	(Optional) If set to <i>yes</i> , then rather than using previously collected, document-wide labelled values, the labelled values are re-examined on each page as it is processed. This can be important if the same label is used on each page with different values, as may be the case for <i>WmlMediaType</i> or <i>WmlMediaPosition</i> . Default: no.

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

Add out-for-mail barcode

This DCC action adds a barcode to a page. The barcode is created from the internal job ID of the document. For example, if the job ID is *157478*, then the barcode will encode the string *J157478*.

Parameters

The parameters for the **Add out-for-mail barcode** action are:

<i>Barcode type</i>	The barcode type to be added. Available barcode types are: <i>code39</i> , <i>usd3</i> , <i>code128</i> , <i>code128auto</i> , <i>msi</i> , <i>msi2mod10</i> , <i>msimod11mod10</i> , <i>msimod10</i> , and <i>msimod11</i> .
<i>bcx</i>	The horizontal position of the barcode location from the left hand edge of the page in PDF units (1/72 inch or 0.353 mm).
<i>bcy</i>	The vertical position of the barcode location from the bottom edge of the page in PDF units (1/72 inch or 0.353 mm).
<i>bch</i>	(Optional) The height of the barcode in PDF units.
<i>bcw</i>	(Optional) The maximum width of the barcode in PDF units. If the calculated width of the barcode is greater than this value then the barcode will be scaled down, but only to a minimum of 80% of the calculated width, to preserve readability.
<i>bctext</i>	(Optional) Font size in points of barcode text. If provided, then the text encoded by the barcode will be displayed below the barcode. The text will be shown in Courier font at the given font size.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Apply media controls

This DCC action specifies a set of media tags to be applied to a page or document, in order to select a paper stock by name or by tray number, and optionally to specify duplex settings.

Parameters

The parameters for the **Apply media controls** action are:

<i>mediatype</i>	Media type (named paper stock) to be applied.
<i>mediaposition</i>	Media position (tray number) to be applied.
<i>plex</i>	(Optional) Sets <i>simplex</i> or <i>duplex</i> for the given page or document. Default: <i>unset</i> .
<i>tumble</i>	(Optional) If set to <i>yes</i> , activates the duplex tumble mode. Default: <i>unset</i> .
<i>targetpage</i>	(Optional) If specified, sets the target page. If omitted, the current page will be used when this action is applied to a specific page, and the first page will be used when it is applied to a complete document.
<i>replaceexisting</i>	(Optional) If set to <i>yes</i> , then any existing media tags are superseded by the new tags. Default: <i>no</i> .

Note

One of *mediatype* or *mediaposition* should be specified, but not both.

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

If applied to a document, then the media tags are applied at the start of the document, but not on each page.

Associated pages on logic

This DCC action uses logic to determine whether certain page types should be present in the document.

For example, it may be necessary to ensure that a certificate is present if the document type is “sales” and a specified field on the letter evaluates as “C”.

Parameters

The parameters for the **Associated pages on logic** action are:

Logic expression	See “Logic-based DCC actions”.
<i>requiredpage</i>	(Optional) A page type which is required. Any number of <i>requiredpage</i> parameters may be supplied.
<i>require</i>	(Optional) Whether <i>all</i> or <i>any</i> required page types must be present. If set to <i>any</i> , then the presence of any one of the <i>requiredpage</i> types will be sufficient to allow the document to pass. Default: <i>all</i> .
<i>forbiddenpage</i>	(Optional) A page type which will cause the document to be held if present. Any number of <i>forbiddenpage</i> parameters may be supplied.

Application

This action can be applied to an entire document, whether the document's type has been determined or not.

Check address

This DCC action checks that the address on a letter is found and contains a minimum number of lines and a valid postcode.

The validity check of the postcode extends to ensuring that its format is legal, but it does not check that the postcode matches the other components of the address.

Parameters

The single parameter of this DCC action is the minimum number of lines that the address should contain, including the postcode and the addressee name. The default is 4 lines.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Check date range

This DCC action ensures that a date, found as a labelled value, lies within a range of days with respect to the current system date.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

Parameters

The parameters for the **Check date range** action are:

<i>Label</i>	The text label for the date to be checked.
<i>past range</i>	Check that the labelled date is no further in the past than this number of days.
<i>future range</i>	Check that the labelled date is no further in the future than this number of days.
<i>notbefore</i>	(Optional) If the labelled date is before this date, specified as YYYY/MM/DD, then the document will be held.
<i>notafter</i>	(Optional) If the labelled date is after this date, specified as YYYY/MM/DD, then the document will be held.
<i>absentlabelok</i>	(Optional) Whether a missing label is acceptable. If <i>no</i> , the document is held if the specified label is not found. Default: <i>no</i> .

The remaining, optional parameters for this action are used to refine the process of selecting the labelled value, and are as detailed for the **Get labelled value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Check monetary value

This DCC action checks the value associated with a label to determine if it is a monetary value (determined by the presence of a currency symbol), and if it is within a user-specified range.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

If the specified value is found to be less than the minimum or greater than the maximum (if given) then the document will be held for review.

Parameters

The parameters for the **Check monetary value** rule are:

<i>Label</i>	The value to be checked.
--------------	--------------------------

<i>min</i>	The minimum acceptable value.
<i>max</i>	(Optional) The maximum acceptable value.
<i>absent label ok</i>	(Optional) Whether a missing label is acceptable. If <i>no</i> , the document is held if the specified label is not found. Default: <i>no</i> .

The remaining, optional parameters for this action are used to refine the process of selecting the labelled value, and are as detailed for the **Get labelled value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Check numeric value

This DCC action checks the value associated with a label to determine if it is a numeric value within a user-specified range.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

If the specified value is found to be less than the minimum or greater than the maximum (if given) then the document will be held for review.

Parameters

The parameters for the **Check numeric value** action are as detailed for the **Check monetary value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Check salutation and address

This DCC action checks the salutation (“Dear Mr...”) and address on a letter. It checks that the address contains a minimum number of lines and a valid postcode, that the name used in the salutation matches the addressee, and that the capitalisation in the salutation is correct.

If the salutation is impersonal (“Dear Sir or Madam”, “Dear Sirs” and so on), the salutation/addressee check is not made.

The address and addressee are stored into the document and the DCC database as `recipientAddress` and `recipient`.

The validity check of the postcode extends to ensuring that its format is legal, but it does not check that the postcode matches the other components of the address.

Incorrect capitalisation can optionally be corrected automatically, but if there is found to be a mismatch between the name used in the salutation and the addressee, then the document will be held for review.

Parameters

The parameters for the **Check salutation and address** action are:

lines The minimum number of lines that the address should contain, including the postcode and the addressee name. Set to 0 to disable the address check (i.e. check salutation only). Default: 4.

fix If set to *yes*, then mistakes in the capitalisation of the name are fixed. Default: *yes*.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Detect phrase flags

This DCC action searches for a number of phrases on the documents and stores them as a *flagset*. The phrases found are stored as a comma-separated list against the specified flagset name in the DCC database, and, optionally, in the PDF document information dictionary.

For example, if the flagset name was specified as “underwriter”, the possible phrases were “Insurer A”, “Insurer B” and “Insurer C”, and only “Insurer A” and “Insurer B” were found, then the list stored to the database and, optionally, to the PDF document would be:

```
underwriter=Insurer A,Insurer B
```

Parameters

The parameters for the **Detect phrase flags** action are:

<i>Name</i>	The name for the group of phrase flags.
<i>Possible phrase 1</i> , [<i>Possible phrase 2...</i>]	Phrases to be included in the group. Any number of possible phrases may be supplied.
<i>addtodoc</i>	(<i>Optional</i>) Whether to add to the PDF document information dictionary. Default: <i>no</i> .
<i>matchcase</i>	(<i>Optional</i>) Whether to match the case of the given possible phrases. Default: <i>no</i> .

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Disallow phrases

This DCC action allows the user to specify a list of phrases which, if found, should cause the document to be held.

Parameters

The parameters for the **Disallow phrases** action are:

Phrase 1, Phrases which, if found, should cause the document to be held. Any
[*Phrase 2...*] number of phrases may be supplied.
matchcase (*Optional*) Whether to match the case of the given phrases. Default: *no*.

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

Electronic document delivery

This DCC action gathers the details required for electronic delivery of a document, and stores them to the PDF for use in the electronic delivery stage.

This action depends on data extracted from the document using page type actions such as **Get labelled value** or **Get positioned value**.

Parameters

The parameters for the **Electronic document delivery** action are:

Email Stored label for the email address. This is either the label, or the value of the *storedas* parameter, for the email address labelled or positioned value.
firstname (*Optional*) Stored label for the first name. Required for delivery via web portal unless a salutation check is applied.
lastname (*Optional*) Stored label for the last name. Required for delivery via web portal unless a salutation check is applied.
mobile (*Optional*) Stored label for the mobile number. Required for delivery notifications by SMS.

Application

This action can be applied to an entire document, whether the document's type has been determined or not.

Ensure field is filled

This action DCC ensures that a value is present for a given label. If the value for the label is missing, the document is held.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

Parameters

The parameters for the **Ensure field is filled** action are as detailed for the **Get labelled value** action, with the addition of the following:

absentlabelok (*Optional*) Whether a missing label is acceptable. If *no*, the document is held if the specified label is not found. Default: *no*.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Get labelled document reference

This DCC action takes the value for a given label and stores it as the unique document reference in the DCC database. If the value for the label is missing and *absent labelok=no*, the document is held.

Note

The document reference found by this action overwrites any previously stored document reference.

An additional feature of this action enables the document reference to be printed in barcode format on the document.

This action is based on finding a *labelled value*. The label should be a normal piece of text located on the page. It is permissible for the label and value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

Parameters

The parameters for the **Get labelled document reference** action are as detailed for the **Get labelled value** action, with the addition of the following:

absent labelok (Optional) Whether a missing label is acceptable. If *no*, the document is held if the specified label is not found. Default: *no*.

The remaining, optional parameters for this action are used to specify barcode parameters for printing the document reference on the page in barcode form, and are as detailed for the **Add out-for-mail barcode** action. At least *barcode*, *bcx*, and *bcy* should be specified.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Get labelled value

This action finds a labelled value and stores it in the DCC database using either the label itself as key, or the value of the *storeas* parameter, if given.

The label should be a normal piece of text located on the page. It is permissible for the labelled value to be located outside the printable area of the page, so that it may be acted upon, but not seen when viewing the document.

The value associated with the label should be located near to the label. For example, in Figure 9.6, the label "Annual Premium" can be used to indicate the value £284.35, while "Total Amount Due" can be used to indicate the value £334.35.

Figure 9.6. Example of labelled values in a document

Annual Premium	£284.35
Legal Assistance:	£35.00
Travel Accident Plan:	£15.00
Total Amount Due:	
£334.35	

The parameters to this action are used to refine the location of the value with respect to the label.

All user-defined labelled values are stored in the DCC database field `labelledValues`. This field exists both in the table `docTable`, which contains information about held documents, and in `outputDocTable`, which stores information about output documents. If `addtodoc=yes`, the value is also stored to the PDF document information dictionary.

Parameters

The parameters for the **Get labelled value** action are:

<i>Label</i>	The text label for the value of interest.
<i>butnot</i>	(Optional) A text token that should <i>not</i> appear in the same phrase as the label.
<i>end</i>	(Optional) A text token which marks the end of the value, for example “:”.
<i>xrange</i>	(Optional) Horizontal search range in PDF units (1/72 inch or 0.353 mm). Default: 10000.
<i>yrange</i>	(Optional) Vertical search range in PDF units (1/72 inch or 0.353 mm). Default: 10000.
<i>xfuzz</i>	(Optional) Horizontal search “fuzz” (PDF units). Default: 36.
<i>yfuzz</i>	(Optional) Vertical search “fuzz” (PDF units). Default: 36.
<i>storeas</i>	(Optional) Storage label (if required to be different from <i>Label</i>).
<i>storeemptyvalue</i>	(Optional) Whether to store an empty value. Default: <i>no</i> .
<i>addtodoc</i>	(Optional) Whether to add the labelled value to the PDF document information dictionary. Default: <i>no</i> .

The optional *butnot* and *end* parameters allow refinement of how the label is selected. *end* defines a separator between the label and value text, which is useful if the label and value appear in a single text element. *butnot* specifies text that should *not* appear in the same phrase as the label, making it possible to avoid matching the label text incorrectly in certain circumstances.

For example, consider a document in which “your premium” was being used to check a monetary value which appeared in a side column of the covering letter. If the body of the

letter contained the sentence “We have pleasure in confirming that your premium covers you...” then the *butnot* parameter could be set to *covers* to ensure that the value in the side pane was always correctly located and the value in the body ignored.

The parameters *xrange*, *yrange*, *xfuzz*, and *yfuzz* determine the expected placement of the value with respect to the label.

The value associated with the label is usually to the right of or below the label, but can be to the left or above. The *xrange* parameter defines the range (in PDF units of 1/72 inch or 0.353 mm) within which the value is expected to be found to the right of the label. The nearest candidate value within this range is chosen. The *yrange* parameter specifies the distance in PDF units within which the value is expected to be found below the label.

To search for a value to the left of the label, specify a negative value for *xrange*. To search above the label, specify a negative value for *yrange*.

The order in which a value is sought is as follows: first, a value is searched for within the same text element as the label (unless *xrange* is set to 0). Second, a value is searched for amongst text elements to the right (or left) of the label, according to the value of *xrange*. If no value is found in this range, or if *xrange* is set to 0, then a value is sought below (or above) the label, according to the value of *yrange*.

If a value is present to the right of, but not exactly in line with, the label, the parameter *yfuzz* can be used to specify a vertical range which is acceptable for the value's vertical position. For example, if *xrange=72* and *yfuzz=18*, then the value for a label should appear within one inch to the right of the label, and in a band extending a quarter of an inch above and below the baseline of the label text. The *xfuzz* parameter performs a similar function for searching for values which appear above or below the label.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Get positioned value

This DCC action is similar to **Get labelled value**, but instead of using a text label as the starting point from which the value should be found, a user-specified position is used. The starting position is typically the expected co-ordinates of the value itself, in PDF units (1/72 inch or 0.373 mm) from the top left of the page.

As with labelled values, user-defined positioned values are stored in the DCC database field *labelledValues*, with the key specified by the *storeas* parameter, since there is no label in this case. If *addtodoc=yes*, the value is also stored to the PDF document information dictionary.

Parameters

The parameters for the **Get positioned value** action are:

xpos The horizontal position from the left hand edge of the page in PDF units (1/72 inch or 0.353 mm).

<i>ypos</i>	The vertical position from the top edge of the page in PDF units (1/72 inch or 0.353 mm).
<i>storeas</i>	(Optional) Storage label. This must be specified for the value to be stored to the DCC database.
<i>storeemptyvalue</i>	(Optional) Whether to store an empty value. Default: <i>no</i> .
<i>addtodoc</i>	(Optional) Whether to add the labelled value to the PDF document information dictionary. Default: <i>no</i> .

The remaining, optional parameters for this action, *xrange*, *yrange*, *xfuzz*, and *yfuzz*, are used to refine the position of the value with reference to the starting position, and are as detailed for the **Get labelled value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Get value array

This DCC action gets a positioned array of values and stores it to the DCC database. It can be used to retrieve a short text containing several values, for example “1234/MR/39/Q/R”. The user must specify the character used to separate distinct values, and also the labels that should be used to store each element in the DCC database. This is done by specifying a single parameter *storeas*, in which the labels are given as an array separated by the chosen separator character.

It is possible to store one or more values within the array as the unique document reference, by including the special label *docref* within the *storeas* parameter. If multiple elements are selected to form the reference, the reference will contain those elements separated by the separator character.

As with labelled values, the values in a value array are stored in the DCC database field `labelledValues`, with the keys specified by the *storeas* parameter. If *addtodoc=yes*, the values are also stored to the PDF document information dictionary.

The value array is located in the same way as for the **Get positioned value** action.

Parameters

The parameters for the **Get value array** action are:

<i>xpos</i>	The horizontal position from the left hand edge of the page in PDF units (1/72 inch or 0.353 mm).
<i>ypos</i>	The vertical position from the top edge of the page in PDF units (1/72 inch or 0.353 mm).
<i>separator</i>	The character that separates the values in the array, for example “/”.
<i>storeas</i>	An array of labels, divided by the specified separator character. Use <i>docref</i> to place an element into the document reference, e.g. <i>storeas=docref/docref/label13</i> .

addtodoc (Optional) Whether to add the labelled value to the PDF document information dictionary. Default: *no*.

The remaining, optional parameters for this action, *xrange*, *yrange*, *xfuzz*, and *yfuzz*, are used to refine the position of the value array with reference to the starting position, and are as detailed for the **Get labelled value** action.

Application

This action is a page type action only: it can be triggered for pages of a known type, or for all page types, but it can't be applied to an entire document.

Hold on logic

This DCC action uses logic to determine whether to hold a document.

A logic expression is evaluated, based on the value of detected labelled values, flagsets, page types, document types, or the processing queue name. If the logic evaluates to true then the document is held.

Parameters

The parameters for the **Hold on logic** action are:

Logic expression See "Logic-based DCC actions".

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

Labels must match

This DCC action ensures that a set of labelled values (obtained for particular page types) all match particular values. If the labelled values do not match, the document is held.

Parameters

The parameters for the **Labels must match** action are:

label=matchValue, List of labels paired with the values that should be
[*label2=matchValue2...*] matched.

Application

This action may only be applied to all documents. It makes use of values obtained by page type and document type actions.

Labels must not match

This action ensures that a set of labelled values (obtained for particular page types) do *not* match particular values. If the labelled values match, the document is held.

Parameters

The parameters for the **Labels must not match** action are:

label=nonMatchValue, List of labels paired with the values that should *not*
[label2=nonMatchValue2...] be matched.

Application

This action may only be applied to all documents. It makes use of values obtained by page type and document type actions.

Set output queue

This DCC action sets the output queue to a specified value, unless, optionally, the original output queue is a given value.

Parameters

The parameters for the **Set output queue** action are:

Output queue The new value for the output queue.
butnotifoutqueueis (Optional) If set, the action will check whether the original output queue matches the given value; if it does, the output queue will not be altered.

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

Store PDF size

This DCC action gets the size of the PDF in bytes before any changes are applied, and stores it to the DCC database and to the PDF document information dictionary.

The value is stored in the DCC database field `labelledValues`, with the key `WmLPdfOrigSize`. This field exists both in the table `docTable`, which contains information about held documents, and in `outputDocTable` which stores information about output documents.

The size is stored to the PDF document information dictionary as a separate field, with the key `WmLPdfOrigSize`.

Parameters

This DCC action has no parameters.

Application

This action can be applied either to a page or to an entire document, whether the document's type has been determined or not.

Tag output queue

This DCC action gets the final DCC output queue, after any adjustments made by other DCC actions, and adds it as the value of the `WmLDccOutQ` tag to all pages of the document.

The tag is a text element of the form `WmLDccOutQ=myqueueuename`, placed at a given position on each page.

Parameters

The parameters for the **Tag output queue** action are:

- xpos* (Optional) The horizontal position from the left hand edge of the page in PDF units (1/72 inch or 0.353 mm). Default: 100.
- ypos* (Optional) The vertical position from the bottom edge of the page in PDF units (1/72 inch or 0.353 mm). Default: 0.
- fontSize* (Optional) The font size in points at which to display the tag. Default: 6.

Application

This action can be applied to an entire document, whether the document's type has been determined or not.

Trigger output queue

This sets the output queue based on information extracted from the document and stored to the PDF document information dictionary using page type actions such as **Get labelled value** or **Get positioned value**.

If the mapping between values and queue names is provided and contains the stored value, then the queue name specified for that value will be used. Otherwise, the stored value itself is used for the output queue.

Parameters

The parameters for the **Trigger output queue** action are:

- valueStoredAs* Stored label for the queue value.
- value 1, queue name 1, [value 2, queue name 2...]* (Optional) Mapping between values and queue names, in case the stored values do not match the target queue names.

Application

This action can be applied to an entire document, whether the document's type has been determined or not.

9.3. Testing the configuration

You can test the configuration at any stage by applying it to one or more example documents that you have uploaded to the system.

Example document analysis

The **Example document analysis** section lists all uploaded example documents, and provides controls to upload, refresh, view, and delete example documents.

Click on “Upload” to upload a new example document. Choose a document that is representative of a type of document that you will be sending through the system. When the

document has been uploaded, it will be analysed using the current profile (including any unpublished changes), and the results of the analysis will be made available to view.

You can refresh the analysis of an existing example document by clicking the refresh button (🔄) for that entry. To refresh all existing example documents, click the “Refresh” button at the bottom of the list of example documents.

Note

A full refresh of all example documents may take some time to complete.

A summary of the example document analysis results is given in the list of example documents. The overall document type and the page types identified by the analysis are shown for each document. The expected outcome of sending the document through the system is also shown, that is, whether the document would be released, or held for some reason.

When the currently selected tab is “Document Type Configuration” or “Page Type Configuration”, example documents that identify as or contain the current type are highlighted.

Click on the information icon (ℹ️) next to an entry in the list of example documents to bring up detailed results of the example document analysis.

Example document analysis results

Document analysis summary

The document analysis summary provides an overview of the document analysis results. The identified outcome, document type, and page types are given, together with the document reference, if found, and any user-defined labelled values and flags.

If a document identifier page was found, details of the document identification results can be obtained by clicking on **View scores** next to the document type. Click on **View all pages** to access page type identification results for all pages in the document.

Document identification results

The document identification results section indicates the scores achieved for all document types in the profile which are candidate types for the example document. You can optionally choose to **Show results for all document types**. Click on a document type name to display a count of each document identification phrase for that type. This can help you to determine if the document identification phrases that you have chosen are successful.

The contents of the document identifier page can be viewed by clicking on **View contents**. This gives a list of text elements appearing on the document identifier page(s), together with the position of each element. This can help to determine the exact wording to use in identification phrases, and the appropriate parameters for actions such as **Get labelled value**.

Page identification results

The page identification results section lists the type of each page in the document, as identified by the analysis. Click on the entry for a given page to bring up a comparison of the

scores achieved by that particular page for all page types in the profile which are candidate types for the page, and click on an individual page type to display a count of the page identification phrases for that type. This can help you to refine the page type identification settings.

The contents of each page can be viewed by clicking on **View contents**. This gives a list of text elements appearing on the page, together with the position of each element. This can help to determine the exact wording to use in identification phrases, and the appropriate parameters for actions such as **Get labelled value**.

9.4. Using document content control

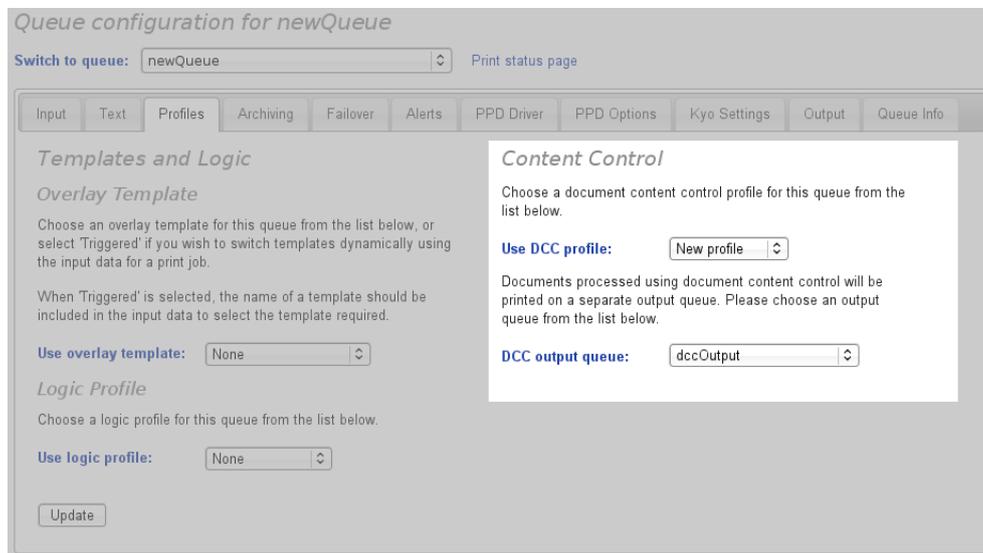
Setting up a queue for document content control

To make use of a DCC profile, it must be bound to a queue, as with templates and logic profiles.

Note

A document content control profile must be *published* (see “Profile status”) to be available for use with a print queue.

Figure 9.7. Binding a DCC profile to a queue



To bind a DCC profile to a queue, go to the queue configuration page at **Print Platform > Queue Config**, choose a queue, and select the **Profiles** tab. Choose a DCC profile from the dropdown menu of published profiles (Figure 9.7).

Documents released by the document content control system will be sent to a separate output queue, so you will be prompted to select an appropriate queue for this purpose.

Press **Update** to apply your changes. Print jobs on the configured queue will now be processed using the selected DCC profile.

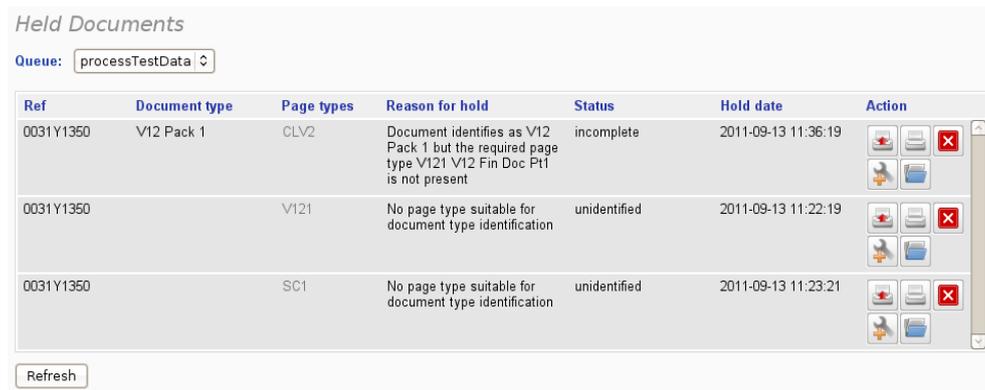
Note

A queue configured for document content control will produce a status page for every job handled by the document content control system, so you may wish to configure its output to be null.

Held documents

The held documents page displays a list of all documents that are currently being held by the document content control system (Figure 9.8).

Figure 9.8. Held documents



The list can be filtered on a per-queue basis, by selecting a queue from the **Queue** dropdown menu, or reloaded using the **Refresh** button.

Each entry gives the document reference, the composition of the document as determined by the system, the reason that it was held, the queue on which it was processed, and the hold date.

The following actions are available for a held document:

-  View the document. The document will be opened for viewing in a new window.
-  Print the document. You will be prompted to select a queue on which to print the document.
-  Delete the document. The document will be removed from the held documents record.

-  Add the document to the configuration. The document will be added to the list of example documents for the profile through which it was processed, so that you can make changes to the configuration of the profile and test the results on this document.
-  Release the document. The document will be printed, and then removed from the held documents record. You will be prompted to select a queue on which to print the document.

Reasons for holding a document

Any documents that are deemed by the system to be incomplete or otherwise unfit for release will be held back for review by system administrators. Some of the reasons for holding a document are listed below.

Document not identified

An *unidentified* document is one which cannot be identified by the system. This may occur if the document is not found to contain an identifier page (which may occur if the identifier page was not recognised by the system), or if there is not enough information to identify the document type from the contents of the identifier page.

Documents which do not include an identifier page may be completed by the addition of another document portion containing an identifier page, as long as there is a unique document reference that occurs on each separate part, providing a means of identifying the separate sections as part of the same document.

If the document identifier page is present but has not been recognised, or if it has not been possible to identify the document type from the contents of the identifier page, the configuration will need to be modified as necessary to identify the page or document type before the system will allow the document to be released.

The stipulation to hold unidentified documents may be relaxed using the **Release unidentified documents** option in the profile settings (see “Releasing unidentified documents”).

Incompleteness

An *incomplete* document is one which does not contain all its required constituent parts. Documents with this status may be completed by the addition of one or more sections containing the missing constituents, as long as there is a means of identifying these separate sections as part of the same document, such as a unique reference that occurs on each separate part.

Failed DCC action

A document may be held if a DCC action *fails*. This may mean that the document failed a validation check, or that a manipulation or information-gathering action failed to complete satisfactorily. Documents with this status will be held for review by a system administrator.

Output documents

The output documents page displays a list of all documents that have passed through the document content control system.

By default, documents for the current date are displayed. You can also search for documents that match a particular search string, or display all documents in a specific date range.

Each entry gives the document reference, document type, number of pages, page types (as short tags), as well as information about how the document was handled: whether it was printed directly, collated, or released (the “Status” column), the queues on which it was processed and output, and the release date.

Searching output documents

You can search for documents that contain certain terms within their metadata (the information displayed for each document), or list all documents in a specific date range.

Searching document metadata

Searching document metadata allows you to find particular documents by matching terms occurring in the document metadata. For example, you can enter a document reference to find a specific document, or a page type tag to find any documents including that page type. Search terms can be combined, allowing you to refine the search results as required.

To search for matching output documents, enter search terms in the **Search all documents** search field and click the **Search** button.

Search tips

Use space or comma separated search terms.

You can group terms containing spaces or commas using double or single quotes: *"direct print"* or *'direct print'*.

Searches are case insensitive: *B12345* and *b12345* will return the same results.

You can use the terms *AND* and *OR* (upper or lower case) to determine how to combine search terms. *AND* is the default. The search *Cover letter* is equivalent to *Cover AND letter* and would not return the document type *Cover note* (but *Cover OR letter* would).

Use the following forms to search for a specific date or time: *2011-02-14* or *15:48*.

Searching by date range

You can display documents for a specified date range using the fields in the **Select report period** box.

To list output documents for a given date range, choose the required period (day, week, month, or custom) and enter a start date (and, for custom date ranges, an end date). Click on the **Submit** button to return the results.

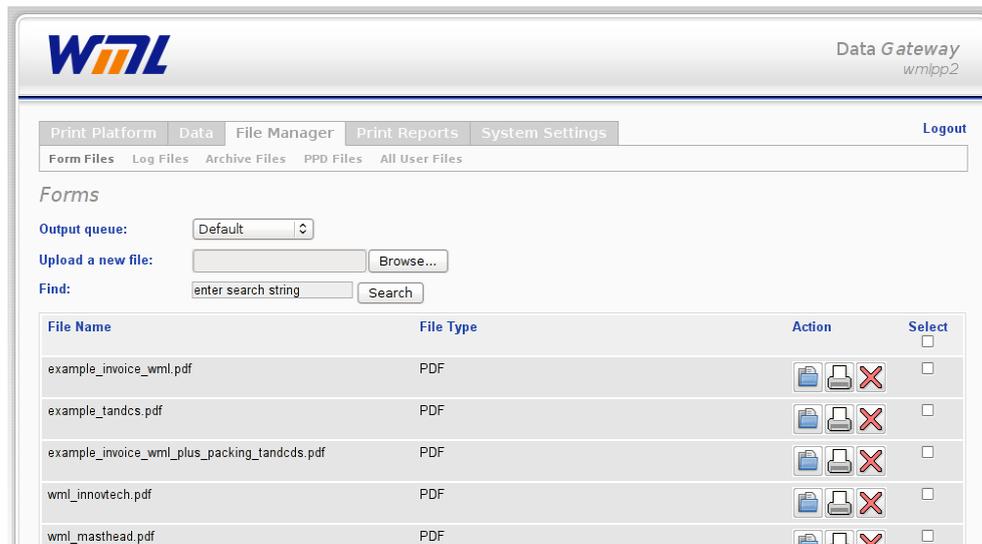
Note

Dates should be of the form *DD/MM/YYYY* or *YYYY-MM-DD*, where any non-numeric separator character is allowed, or *YYYYMMDD*, without separators. A separator character is required if the date order is day, month, year.

Chapter 10. File Manager

To access the built in file manager, select **File Manager** from the top level menu. Your browser will display a page similar to Figure 10.1.

Figure 10.1. The DataGateway's file manager



The File Manager has a number of sections listing files relating to different features of the DataGateway.

- Form Files: User-uploaded PDF files for use as forms or digital stock.
- Log Files: Daily system logs stored on a network share or on internal storage.
- Archive Files: Archived print jobs, in PDF format, stored on a network share or on internal storage.
- PPD Files: User-uploaded PPD driver files.
- All User Files: Shows PDF form files and any files generated with the print diagnostics feature (enabled in **Print Platform > Queue Config > Output**).

To upload a file to the DataGateway, use the “Upload a new file” field.

Each page of the File Manager contains a search field which will display files whose name matches the entered search term.

Where relevant, listed files can be viewed, printed and deleted from the system using the icon buttons in the “Action” column. When an option to print is given, the output queue to which the files will be sent is chosen from the “Output queue” menu.

10.1. Form Files

The file browser shows a list of the files stored on the internal flash or hard disk of the DataGateway. Note that not all files stored on the DataGateway are displayed. Archived files and log files are not displayed here; the list is intended for user-supplied files such as PDF overlays and printer macros, and diagnostic files which may be generated by the DataGateway job-processing software.

10.2. Log Files

Shows daily system logs stored either on a network share, or on internal storage. System log settings are configured in **System Settings > System Log**.

10.3. Archive Files

Shows files archived to a network share or internal storage. Choose the queue to list archived files for in the menu “List files for queue”. The number of files shown can be limited with the menu “Max files to display”.

10.4. PPD Files

Shows PPD files and gzipped PPD files which have been uploaded to the DataGateway by the user.

10.5. All User Files

This page displays PDF form files and other user-uploaded files. Files generated with the print diagnostics feature (enabled in **Print Platform > Queue Config > Output**) appear here.

Print Diagnostics

For diagnostic purposes, you can have the print job saved before and after it is processed by the DataGateway software. The input data will appear in the file list as `prefilter_queue.bin` and the output as `postfilter_queue.bin`.

This feature has the potential to fill the storage of the device and lead to failed print jobs, so care must be taken with its use. Because of this, the feature will always be switched off following a reboot of the DataGateway.

Enable print queue diagnostics in the Output tab of **Print Platform > Queue Config**.

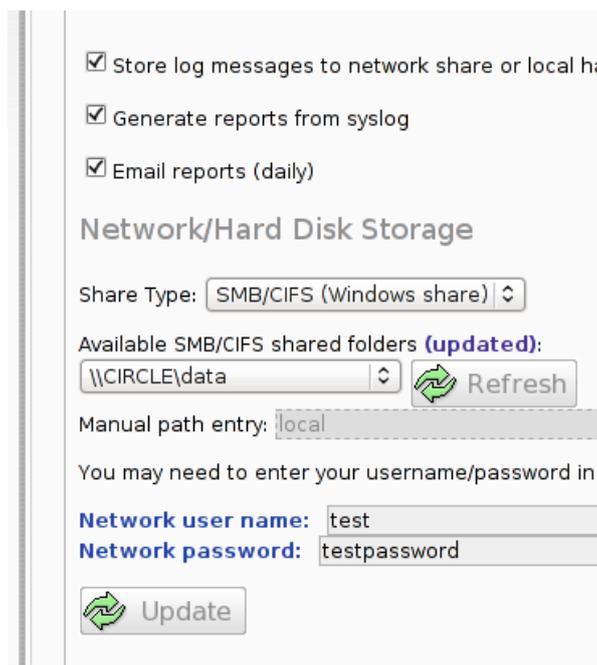
Chapter 11. Print Reports

The DataGateway can provide an audit trail showing what time each job is received by the device, and what time it is sent on to the printer. Print reports are generated from analysis of system log data. The reports can be viewed via the web interface and emailed daily to a group of recipients.

Because the reports are generated from system log data, it is necessary to configure the storage of system logs either to a network share, or to local hard disk storage. Print reports cannot be generated from the volatile system log which is always available from RAM.

Refer to the section called “System logging” in Chapter 5 for instruction on configuring the system log settings. Ensure that **Store log messages to network share or local hard drive** is checked, as shown in Figure 11.1, so that the system log will be permanently stored. The system log will be rotated daily in this case, creating a job log file and a system log file each day, as described in the section called “System logging” in Chapter 5.

Figure 11.1. Syslog settings necessary for report generation



The screenshot shows a web interface for configuring syslog settings. At the top, three checkboxes are checked: "Store log messages to network share or local hard drive", "Generate reports from syslog", and "Email reports (daily)". Below these is a section titled "Network/Hard Disk Storage". It includes a "Share Type" dropdown menu set to "SMB/CIFS (Windows share)". Underneath, it says "Available SMB/CIFS shared folders (updated):" followed by a dropdown menu showing "\\CIRCLE\data" and a "Refresh" button with a circular arrow icon. Below that is a "Manual path entry:" field containing "local". A note states "You may need to enter your username/password in". There are two input fields: "Network user name:" with the value "test" and "Network password:" with the value "testpassword". At the bottom of this section is an "Update" button with a circular arrow icon.

Checking **Generate reports from syslog** will ensure that reports are generated from the daily log files. Reports are generated each morning by processing the YYYY-MM-DD-wmlpp_job.log file. The report is output as YYYY-MM-DD-wmlpp_report.csv. Note that the report date will be one day earlier than the job log date. 2009-07-12-wmlpp_job.log contains the job logs for 2009-07-11 and is used to generate 2009-07-11-wmlpp_report.csv, which is the report for 2009-07-11.

If **Email reports (daily)** is checked, the YYYY-MM-DD-wmlpp_report.csv report will be emailed as an attachment each day, along with a summary of the day's printing. An example is shown in Example 11.1.

Example 11.1. Example email report

```
From: wmlpp@wmltd.co.uk
To: recipient1@example.com, recipient2@example.com
Subject: Report: 2009-07-02-wmlpp_report.csv (from wmlexample2)
Date: Fri, 03 Jul 2009 06:04:33 +0100
```

WMLPP Printing Report: 2009-07-02-tempest_report.csv

Summary for wmlexample2:

```
Total jobs received:          1043
Batch_A:                      179
Batch_B:                      303
raw_input:                    550
Unmatched jobs:              11
Total jobs successfully transferred: 1029
Total jobs not yet transferred: 14
Total Megabytes transferred:  821.559
Total simplex sides received for printing: 29842
Total sheets if printed duplex: 14921
```

A detailed report is attached to this email.

Thank you for using the WML Print Platform.

Attachment: CSV document attachment (2009-07-02-tempest_report.csv)

The format of the report is comma-separated values, with each field enclosed in inverted commas and separated by a comma. The summary information is presented first, followed by a detailed listing of each job. An example is shown in Example 11.2.

Example 11.2. Example report attachment

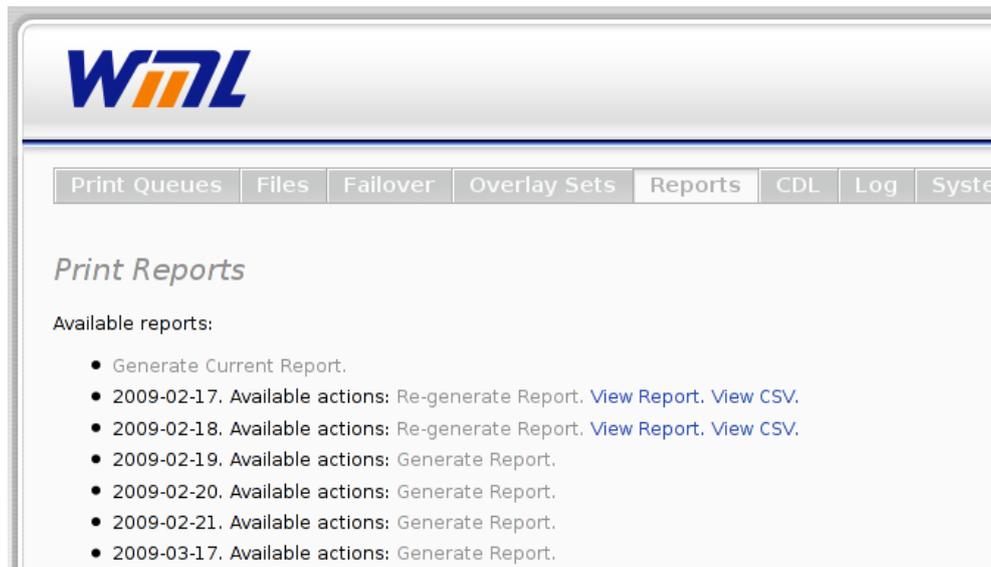
```
"Summary of jobs sent by wmlpp1:"
"Total Jobs","Un-matched","Total MB","Total Sides Printed","Total Duplex\
Pages","Total Completed","Total Not Yet Complete"
"13","13","0.173437","24","16","4","9"

"Detailed listing of jobs sent by wmlpp1:"
"Job Title","Bytes","ID","Simplex Sheets","Duplex Sheets","Received at \
wmlpp1","Complete Send to Print Device"
"qlpd_job","128","1","3","2","14/07/2009 10:21:26","14/07/2009 14:27:39"
```

11.1. Viewing reports in the web interface

To access reports in the web interface, log in and select **Print Reports**. Your browser will display a page similar to Figure 11.2

Figure 11.2. View of available reports



Available reports are shown in a list. The very first entry is **Generate Current Report**. This will generate an up-to-the-minute report based on the current print log. If you click the link, the print log will be processed and then two new links will appear. **View Report** and **View CSV**. Click on either to view the report.

For each dated job log file which has been created in the system log store, an entry is seen on the Print Reports page. Here, you have the option to generate, re-generate or view the report. If you generate a report, a second counter will count the time as the report is generated.

Note

For a production system printing hundreds or thousands of jobs per day, report generation may take as long as a few minutes to complete.

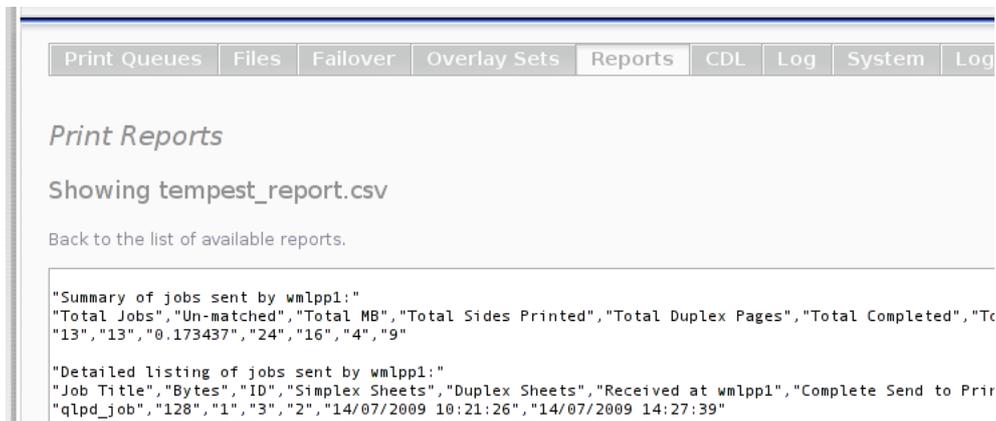
Figure 11.3. Viewing a report



Figure 11.3 shows an example of what you will see if you click **View Report**. You can return to the list of reports by clicking the link **Back to the list of available reports**. The listing of jobs is presented as a table, containing the title, its size, the internal job ID and the number of pages, along with the time the job was received and when it was sent on to the printer.

Figure 11.4 shows the same example data, but on the CSV page, where the data is presented in comma-separated format.

Figure 11.4. Accessing report in CSV format in the web interface



11.2. Reports with a pair of DataGateway devices

The WML DataGateway is able to work in pairs, in which data is transferred from one DataGateway to the other, ensuring secure, audited, automated data transfer across a wide area link.

Note

This feature is not available on other WML products, such as PdfPlus and PrePrint.

The log messages from each DataGateway are stored on both systems. An SSL tunnel is used to transfer the messages in real-time. The ports used by the SSL tunnel are configured in **System Settings > System Logging. Send syslog messages to** should be set to protocol TCP/SSL and the suggested port to use is **5140**. **Receive network syslog messages** should have protocol set to TCP/SSL, and again it is suggested that port **5140** is used. This port will need to be opened on any firewalls which are configured between the DataGateways.

In addition to the real-time log messages, an SSL based secure copy process is used to transfer some files (especially logs recording barcode scans of documents). To enable this transfer, the sending DataGateway should be permitted to make a connection on TCP port **7522** to the receiving DataGateway.

Chapter 12. Appliance Manager

Appliance Manager is an optional module for DataGateway which allows you to manage a group of WML appliances. The module is enabled (with a licence key) on a single DataGateway which becomes the master appliance manager on the network.

Appliance Manager has the following features:

- Catalog individual appliances and create appliance groups in a user-friendly web interface.
- View realtime status information for all appliances.
- Push appliance settings out to an appliance or a group of appliances.
- Apply firmware upgrades to an individual appliance or to a group.
- Push out a reboot or shutdown command to an appliance or group.
- Review managed appliances in a CSV-formatted report.

A time window can be set for all actions to ensure that they are carried out during off-peak hours.

All communication channels between appliance manager and client devices are 256-bit SSL encrypted ensuring that Appliance Manager can be used safely across a public wide-area network.

12.1. The Appliance Manager “push”

Appliance Manager provides the concept of a **push**. A push is the specification of a **configuration**, along with an appliance or group of appliances and a time window during which the push should be applied.

A configuration may be a **settings configuration**, a **firmware upgrade** or a **command**.

A settings configuration is a **backup.tar** file, as generated within **System Settings -> Backups**. It may contain a set of templates or print queues. The settings which can be incorporated into the configuration are listed in the section called “System settings”.

A firmware upgrade file is a standard WML upgrade file, as provided for an individual appliance. A typical firmware upgrade file might have the filename **wmlpp-dg-kvm-sata-tree5-20110125_1142.img**.

Commands which can be distributed with a push currently include only the **reboot** and **shutdown** commands.

A managed appliance will periodically query the appliance manager to see if there are any pending pushes which it should apply. The manager then transfers the data (upgrade image or backup file) and waits for the appliance to apply the update and confirm success.

The main activity for the administrator using Appliance Manager is to create and manage pushes.

12.2. The user interface

Appliance Manager has three sections to its interface:

- **Appliances and Groups:** The main list of appliances and appliance groups;
- **Stored Configurations:** A page to manage uploaded settings configurations (backup.tar files from individual appliances);
- **Stored Firmware:** A page to manage uploaded firmware upgrade files.

12.3. Appliances and groups

This section covers the most important part of the Appliance Manager user interface; **Appliances and Groups**.

Figure 12.1. The Appliances and Groups section of the Appliance Manager.

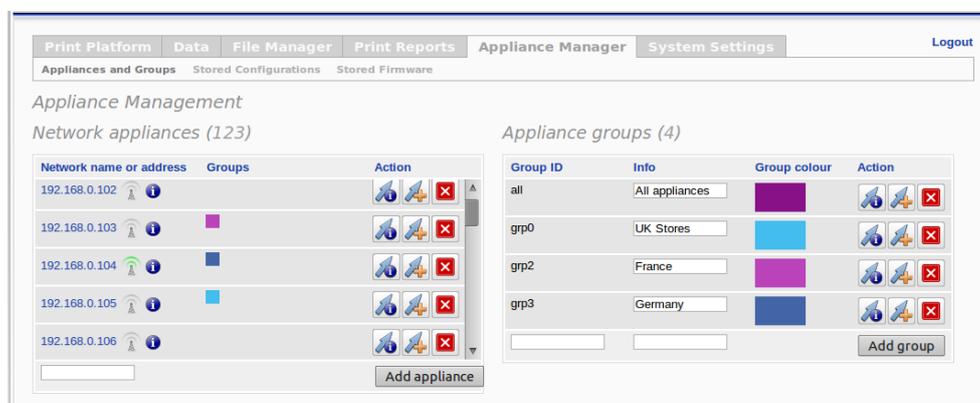


Figure 12.1 shows the **Appliances and Groups** page. To the left is a table headed **Network appliances** listing each appliance being managed. This list is created by the administrator, either by adding entries with the **Add Appliance** row of the table, or by uploading a CSV file containing a list of appliance addresses.

Note

The number of appliances and groups under management is shown in brackets after the table heading.

The **Network appliances** table displays the appliance address (IP address or network name), its online/offline status (🟡), an appliance status button (🔵), group memberships (denoted by coloured blocks) and a set of action buttons (🔵, 🟡 and 🔴).

Note

The “online status” icon will be green (🟢) if the managed appliance is online and successfully sending status messages to the appliance manager.

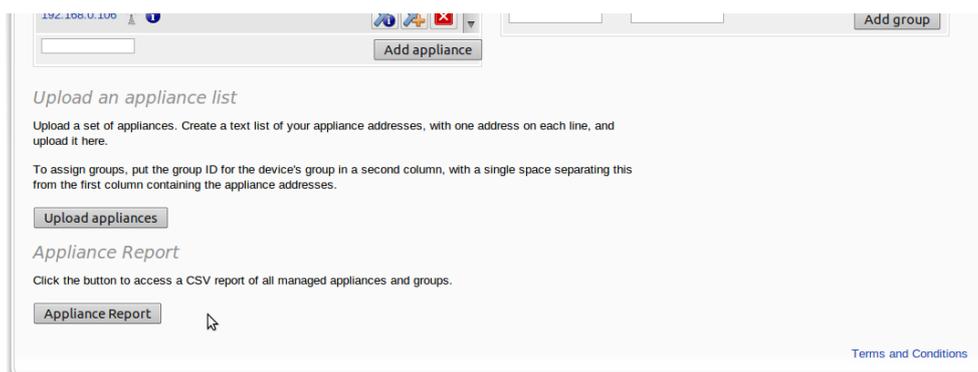
If the managed appliance is not sending status messages (even if it is otherwise accessible, for example by ping or the web-interface) then the icon will appear grey ().

The **Appliance groups** table to the right allows the administrator to create groups. A group is given a name (**Group ID**) and a short description (**Info**). It receives an auto-generated colour which is displayed as a large block (**Group colour**). The action buttons in the **Appliance groups** table have the same meaning as those in the **Network appliances** list.

To make an appliance a member of a group, drag the appliance row over and drop it on the group row. The group colour will appear as a small block in the **Groups** column on the **Network appliances** list. To remove an appliance from the group, click on the group block (when you hover over the group block, a white cross will appear).

Populating the appliance list

Figure 12.2. Upload an externally generated list of appliances, or access an appliance report.



Towards the bottom of the Appliances and Groups page, there are two additional sections, shown in Figure 12.2. **Upload an appliance list** provides an interface to use a text file to populate the **Network appliances** list. **Appliance Report** generates a full report of the system and is covered later in this section.

This list of addresses can be a single column of IP addresses or network names as follows in Example 12.1:

Example 12.1. Appliance list (with no groups specified)

```
192.168.0.1
192.168.0.2
192.168.0.3
appliance1.local.lan
appliance2.local.lan
appliance3.local.lan
```

Addresses on this list are added to the existing list of network appliances without first deleting any pre-existing entries. If any of the appliances are already listed then duplicates are *not* added.

Example 12.2. Appliance list with group specified in column two

```
192.168.0.1 London
192.168.0.2 London
192.168.0.3 London
appliance1.local.lan London
appliance2.local.lan Reading
appliance3.local.lan Reading
```

To set group membership when uploading the list of appliances, include a second column of the appliance group, as shown in Example 12.2. The columns should be separated with a space. Column two is the group ID and should contain no spaces.

Note

If the group specified in the uploaded file does not exist, it will be created.

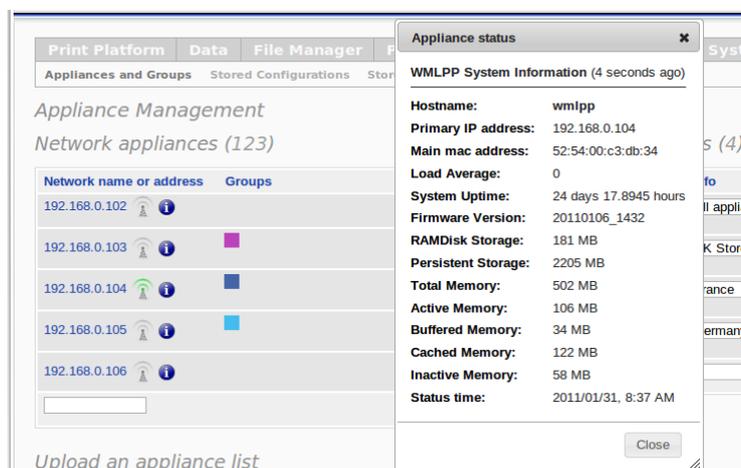
If any groups *are* created, it will be necessary to re-load the **Appliances and Groups** page to display them, as they are not automatically added to the **Appliance groups** table.

Appliance status

To view the status of a managed appliance, click on the status icon (🔍). The status dialog will be displayed, as seen in Figure 12.3.

Note

It will take up to five minutes for the system status to appear for a newly managed appliance.

Figure 12.3. Viewing the status of a managed appliance.

Most of the status fields displayed in **Appliance status** are self-explanatory, but note:

- The **Status time** displayed is the time as reported by the clock on the managed appliance.
- Of the memory fields, the most important is **Active memory**, which should be less than **Total memory** by a sensible margin. (The memory data is extracted from the Linux kernel's `/proc/meminfo`.)

When using Appliance Manager, it is important to ensure that all appliances have the same system time. The best way to achieve this is to make use of the NTP feature (See the section called “System date and time”).

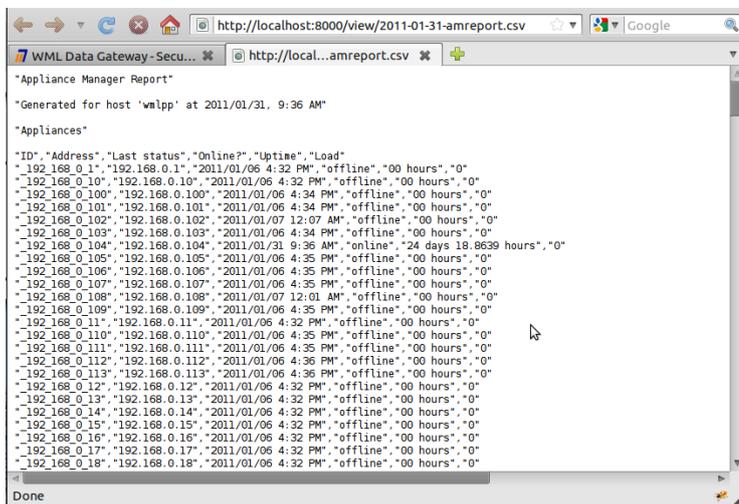
Note

If the system time on the managed appliances differs by a significant amount from the appliance manager, then the SSL-encrypted communication between appliance and manager will fail. (SSL certificates do not validate if they appear to be too old, or exist in the future).

Appliance report

To view a full report, use the **Appliance Report** button. This will generate a comma separated value-formatted report containing:

- All managed appliances, with status
- Groups with lists of appliance members
- Full listing of Pushes, both complete and incomplete, including appliances to which the pushes apply.

Figure 12.4. Viewing the full report of managed appliances and groups.

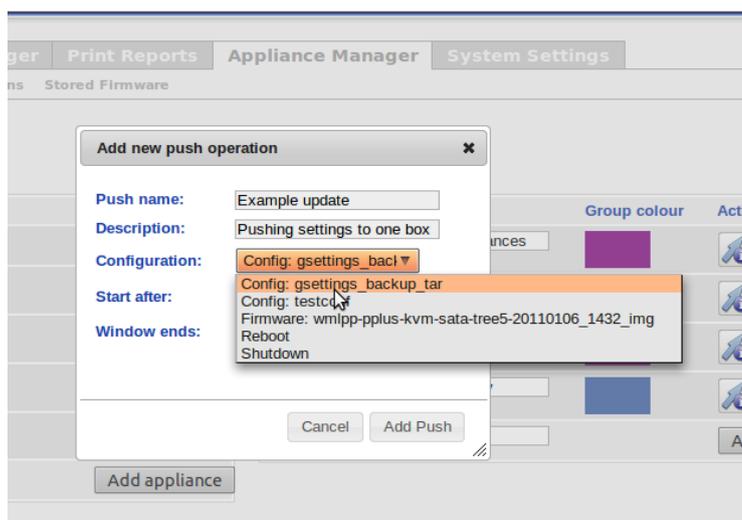
The report will pop-up in a new tab or window in your browser (as in Figure 12.4). Save this file and then import it into a spreadsheet to access and manipulate the information.

Note

You will need to set your web browser to allow pop-up windows for the appliance manager. If you do not do this, you will probably see a message stating that the browser blocked a pop-up when you try to create the appliance report.

Managing pushes

Figure 12.5. Creating a new “push”, and choosing a configuration to push out to the appliance.

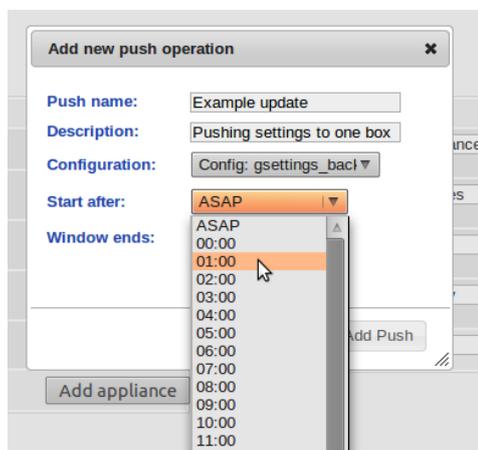


Use the add push icon (📁➕) to create a new push for an appliance or group. A dialog will pop up for the push details, as shown in Figure 12.5. Enter a name for the push, a short description and choose the configuration.

Figure 12.5 shows the configuration menu displaying five options; two example settings configurations, a single firmware upgrade file and the standard commands **Reboot** and **Shutdown**. The number of options in this menu depends on the number of configurations and firmware images which have been uploaded on the **Stored Configurations** and **Stored Firmware** pages.

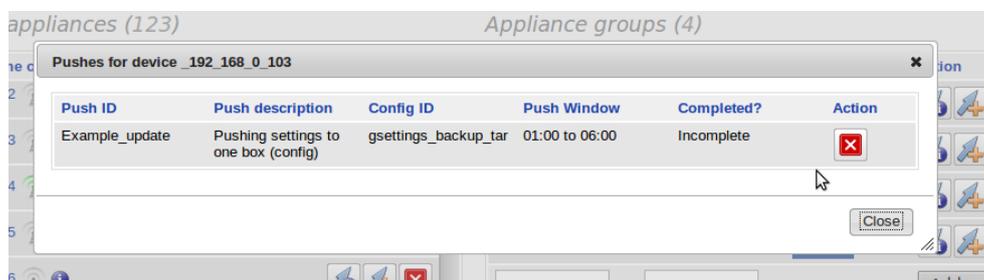
If the push should be applied as soon as possible, choose **ASAP** in **Start after** (Figure 12.6). Otherwise, choose a start time and a time when the push window ends. The system will attempt to apply the push during that window. Unfortunately, it's not possible to specify a particular date for a push.

Figure 12.6. Choosing a time window for the new push.



If this is a push for a group, then it will include all of the appliances in the group *at the time the push is created*. If an appliance is added to the group *after* the creation of the push, it will *not participate* in the push. Likewise, if an appliance is removed from the group after the push was created, it *will still have the push applied* (unless the appliance itself has been deleted).

Figure 12.7. The list of appliance “pushes” for the appliance “_192_168_0_104”.



Once the push has been created, it can be viewed with the push query icon () . This will bring up a dialog box as shown in Figure 12.7. This is a simple table showing the pushes for the appliance or group. The ID of the push is shown, along with its description, the config which is to be pushed out, the time window and whether or not it is complete. If the push is complete, the time of completion is shown.

Note

The appliance ID in Figure 12.7 is _192_168_0_104. This is generated from the appliance address, by conditioning the address into a tag which can be used within Appliance Manager's internal XML configuration files, hence the '_' characters replacing '.' characters.

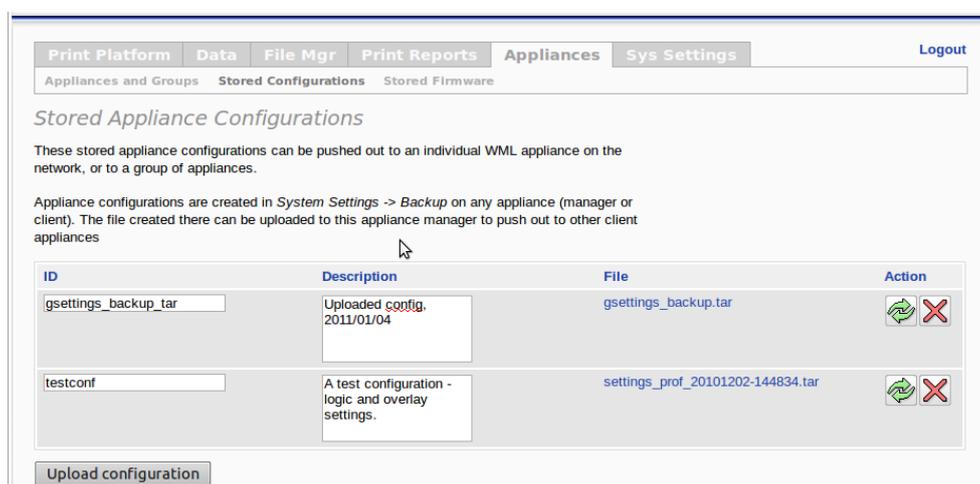
If you hover your mouse over the **Push ID** field, the appliance members of the push are shown. This is useful when reviewing group pushes. Likewise, if you hover over **Completed?** a list of the appliances which have applied the push is shown as well as those which have yet to complete.

Completed pushes are displayed until the administrator decides to delete them. A push can be deleted (✖) whether it is complete or not. If an incomplete push is deleted, those appliances which have not applied the push will stop attempting to do so.

12.4. Stored configurations

This page (Figure 12.8) provides an interface to upload **backup.tar** files for distribution to managed appliances.

Figure 12.8. An interface to upload settings configurations (backup.tar files).



Typically, an administrator would configure a separate appliance with suitable settings. This could be one of the managed appliances, but it does not have to be. He would then save the configuration for the appliance in **System Settings -> Backup**, choosing only the “section” of the settings required. For example, he might save off the templates and logic settings, so that these could be pushed out to the managed appliances without affecting their system settings and print queues.

He uploads the backup.tar file to this interface where it is given an ID based on the file name and a description based on the date. The configuration will now appear in the drop-down menus of the **Add new push operation** dialog in the **Appliances and groups** page.

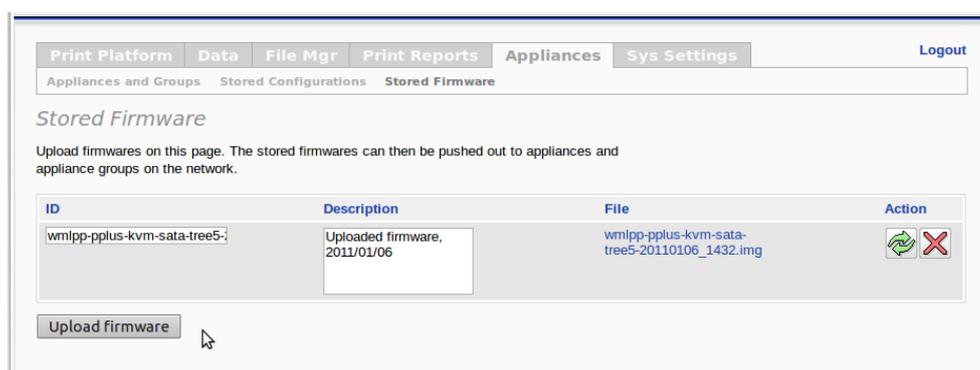
Caution

The interface *will* allow you to delete a configuration which is a member of a pending push. If you do this, then the push will not complete.

The **Action** column provides update and delete buttons to update changes to the **ID** or **Description**, or to delete the entry.

12.5. Stored firmware

Figure 12.9. Upload firmware upgrade files here, ready to be pushed out to managed appliances.



This page (Figure 12.8) provides an interface similar to that in **Stored Configurations**. This is for uploading and managing firmware upgrade files.

The administrator uploads the same firmware upgrade image which he would use to upgrade a single box. Again, it is given an ID based on the file name and a description based on the date. Once uploaded, it is available to be pushed out to upgrade managed appliances.

Caution

The interface *will* allow you to delete a firmware which is a member of a pending push. If you do this, then the push will never complete.

As in **Stored Configurations**, the **Action** column provides update and delete buttons to update changes to the **ID** or **Description**, or to delete the firmware file.

12.6. Appliance Manager for Disaster Recovery

The Appliance Manager system is used to provide a Disaster Recovery (DR) backup system. This allows the firmware, config and data on one DataGateway to be mirrored to a second DataGateway.

To set up a DR pair, designate the master system. Appliance Manager should be licenced on the master system. On the master system, add a single network appliance in **Appliances > Appliances and Groups**. Enter the correct IP address for the DR backup system and press **Add appliance**. Check the **DR** checkbox to designate this as a DR backup system.

All configuration settings and firmware upgrades will now be transferred to the DR backup system. Settings and data (including the contents of the Document Content Control

database) are transferred twice daily. Firmware upgrades are pushed from the main system to the backup system shortly after the main system is upgraded.

Warning

The DR backup system should be running on the same infrastructure as the main system. Either both should be KVM based (such as those which are supplied by WML as hardware appliances) OR both should be VMware based. If one system is KVM based, and the second is VMware based, firmware upgrades applied to the main system will not propagate to the DR backup system.

The DR backup will hold its existing IP address settings until it is “rebooted as a DR main system”. This is done by accessing **System Settings > Status/Control** on the DR backup system. On this system there should be a reboot icon labelled **Disaster Recovery Reboot**. If this is pressed, the DR backup system will reboot and take on the personality of the DR main system, including its IP address settings.

Warning

It is very important to ensure that the main system is offline when the DR back-up is brought up with **Disaster Recovery Reboot**, otherwise IP address conflicts will ensue.

12.7. Appliance Manager Networking

This section describes the ports used by Appliance Manager to communicate with managed devices.

The Appliance Manager **listens** on **UDP port 1554** for incoming connections from managed devices so that these devices can announce their presence. This allows newly added managed devices to be automatically listed in **Appliances > Appliances and Groups**.

The Appliance Manager has a server **listening** on **TCP port 1555** for status messages sent from managed devices. This channel is SSL encrypted.

The Appliance Manager **makes connections** to **TCP port 1556** of managed devices to send firmware, config and command pushes. This channel is SSL encrypted.

Finally, Appliance Managers which have managed DR partners **make connections** to **TCP port 22** on the DR partner(s) to transfer configuration settings. All communications on this channel are SSL encrypted.

Chapter 13. Electronic Document Delivery

DataGateway's *electronic document delivery* system enables delivery of completed documents either by secure document download or by email attachment. This streamlines document fulfilment and provides an audit trail for document delivery.

The electronic document delivery system works by extracting delivery parameters, such as name, email address and mobile number of the intended recipient, from within an incoming document. These details are then used to set up electronic delivery of the document. The document is stored to a system archive, and electronic notification (which may be email, SMS message or both) is sent to the recipient to inform them how they can pick up the document.

If the secure document delivery option is in use, the procedure to pick up the document will involve the recipient logging in to an existing customer system or accessing a secure portal on the DataGateway. The stored document is linked to the recipient's account or details and will be made available for download on successful authentication.

An alternative delivery option is for the document to be delivered by email attachment. This is the simplest method, but it does not offer an audit trail for tracking the delivery of a document.

A key feature of WML's secure document delivery is access to an audit trail for the delivery of the document to the recipient, since all successful or failed attempts to download the document are logged to the system. Repeat notifications can be sent to the recipient if the document remains undelivered, and the system can be set up to fall back to print and mail after a specified time period has elapsed. The administration interface provides controls to resend notification messages and to view or print documents, as well as a document access timeline.

For more information about using DataGateway for electronic document delivery and for access to a live demo of the electronic document delivery system, please contact WML.

13.1. Overview

Electronic document delivery can be applied either to structured input data or to formatted documents of various types. The system extracts delivery parameters, such as the name, email address and mobile number of the intended recipients, from an incoming document. This information is located either by searching for specified keys within the document data or by passing the file through DataGateway's document content control system, which can be configured to search for the required fields and make the information available as document metadata.

The archive feature of DataGateway is used to store documents awaiting electronic delivery. Document details are stored to the electronic documents database, which holds information about each document in the system and which tracks the delivery status, sending of notification messages and so forth.

When a document enters the system, a notification message is sent to the intended recipient. This message can be an email, an SMS message or both, according to the system configuration. The message will inform the recipient about the document and provide information about retrieving the file. The system can be configured to send repeat notification messages if the document remains undelivered.

The document may be delivered either directly, by attachment to the email message sent out by the system, or by secure download. The secure delivery route obliges the recipient to download the document via a secure portal or through an existing customer account, which is associated with the document for download using information such as a document, customer or policy reference. The document is made available for download on successful authentication.

The secure delivery option has the advantage that a document audit trail is produced, with any attempts to download the document being logged to the system. The delivery of a document is considered incomplete until the document has been downloaded successfully by the customer. If a document remains incomplete, repeat notifications may be sent, and the system may be configured to fall back to print incomplete documents for mail after a specified time period.

The electronic delivery customer support interface provides access to the delivery status and access timeline for each document in the system. Notifications can be resent, and documents can be viewed, printed or cancelled. Document records are located by searching on delivery details or document identifiers such as document, customer or policy reference.

13.2. Configuration

The main configuration of the electronic document delivery system is carried out on the **Electronic Delivery > Configuration** page of the DataGateway web user interface, which provides controls to configure core and per-queue settings.

You can also configure the form of the notification messages to be sent out by the system by going to **Electronic Delivery > Message Templates**.

Core and per-queue settings

The core and per-queue settings (Figure 13.1) control the parameters for the electronic document delivery system. Default parameters are defined in the core settings. Several of these can be overridden on a per-queue basis, to allow different queues to be used for different delivery scenarios.

Figure 13.1. Electronic Document Delivery configuration: core and per-queue settings

The screenshot displays two configuration panels. The 'Core settings' panel on the left includes fields for 'Default parameters' (First name key: firstName, Last name key: lastName, Mobile key: mobileNumber, Email key: emailAddress, Customer reference key: customerReference), 'Message preference: Email', 'Email subject line: Your documents are', 'Number of messages to send: 7', 'Hours between messages: 26', 'Revert to hardcopy: [checked]', 'Revert to print after (days): 14', 'Output print queue:', 'SMTP server:', 'SMTP user:', 'SMTP pass:', 'SMS test mode: [unchecked]', 'HTTPS proxy server:', 'HTTPS proxy user:', 'HTTPS proxy pass:', 'Database name: wmlredd', 'Database server: 192.168.0.122', 'Database username: wmlredd', and 'Database password: password'. An 'Update' button is at the bottom. The 'Per-queue settings' panel on the right is titled 'SecureDelivery parameters' and has a dropdown menu set to 'SecureDelivery'. It contains the same parameter fields as the core settings, but with 'Message preference: both SMS and email' (indicated by a print queue icon), 'Email subject line: Your documents are ready f', and 'Output print queue: hp3005' (indicated by a print queue icon). An 'Update' button is also present at the bottom of this panel.

Set up default parameters by editing values in the core settings area. Click on the **Update** button below the core settings to save the changes.

To override a default parameter for a particular queue, select the queue from the dropdown menu above the per-queue settings area. Edit the values as required, then click on the **Update** button underneath the per-queue settings to save the changes. Per-queue parameters that differ from the default parameters are indicated by a print queue icon.

The parameters available for configuration in the core and per-queue settings are described below.

Customer data keys

The **First name key**, **Last name key**, **Mobile key**, **Email key** and **Customer reference key** settings specify the set of keys used to denote those details within the structured input data. If data items with these keys are present in the input data, then the electronic document delivery system will extract the customer details thus defined and store them for use in delivering the output document.

Message configuration

The **Message preference** dropdown controls the form of electronic delivery to be used, and the type of notifications that should be sent. If the message preference is chosen to be "SMS", "Email" or "both SMS and email", then delivery of the document will be via the secure delivery route, with the specified type of notification. If "Email document" is chosen then the document will be sent by email attachment.

Further message configuration settings allow you to specify an **Email subject line** for the emails sent out by the system, the total **Number of messages to send**, and the number of **Hours between messages**. Note that the wording of individual messages can be configured in **Electronic Delivery > Message Templates**.

Revert to print

You can choose to revert to hardcopy if a document has not been delivered after a certain number of days by checking **Revert to hardcopy** and specifying the number of days after which the system should print an undelivered document. **Output print queue** should be set to the name of the queue to be used for printing. This is also the queue used to print when the admin user prints a document from the search page.

SMS settings

You may be required to enter details of the text messaging service that will be used to send SMS messages.

You can check **SMS test mode** to perform system tests which do not actually send an SMS, even if SMS is included in the message preference setting.

SMTP, proxy and database settings

The remaining settings indicate login details for the SMTP server, HTTPS proxy server (if required to gain access to the SMS text service' web server) and electronic document delivery database.

Message templates

You can set up message templates for the notification email and text messages. These templates will be used in order, until the final template is reached. If the system is set up to send any subsequent messages, these will use the final template.

At least one message template needs to be set up for each type of notification in use. You can add messages by going to the "New Email" or "New SMS" tab, entering the message text in the editor area and clicking on the **Save** button. You can delete an existing template by going to the relevant tab and clicking on the **Delete message** button.

Special tags, prefixed with %%, are used to place document/customer identification codes into either text or SMS messages. These will be replaced by the corresponding details when the messages are sent. Together with the customer details, these dynamically inserted details can be used to identify the document on the system that is to be delivered to the recipient. See "Special tags for use in message templates" for a list of the tags which may be included in message templates.

Special tags for use in message templates

- %%custref Document, customer or policy reference
- %%code Access code for secure portal delivery
- %%fname Customer first name for secure portal delivery
- %%lname Customer last name for secure portal delivery

You can test a template by clicking on the **Test message** button. You will be prompted to enter recipient details, then a message will be sent to the recipient using the given template.

Email message templates

Email message templates allow you to specify the wording of customer notification emails. Typically, you might include information about the document and, for secure portal delivery, instructions about how the recipient can gain access to the document. A very simple example is given in Example 13.1.

Example 13.1. Example email template

```
Your document (Ref: %%custref) is ready.  
  
Please visit http://mysite.com/login and log in to your  
customer account to download your document.  
  
Kind regards,  
Sales Team
```

Email messages may be provided in two alternative formats, plain text and HTML. The text version must be provided, while the HTML version is optional. If an HTML version is available, the email message will be sent in HTML format, with the plain version available as a fall-back for mail clients that do not accept HTML emails.

Note

Care should be taken that the text content is the same for both HTML and plain text versions of an email message, since some mail clients may consider messages in which the plain text and HTML parts differ as junk mail.

You can use the HTML template to send a message which includes layout, formatting, images and so on. Standard HTML markup should be used to generate the desired look and feel. For example, Example 13.2 shows an HTML template corresponding to the plain text version in Example 13.1.

Example 13.2. Example HTML email template

```
<p>Your document (Ref: <strong>%%custref</strong>) is ready.</p>

<p>Please visit
<a href="http://mysite.com/login">http://mysite.com/login</a>
and log in to your customer account to download your document.</p>

<p>Kind regards,<br/>
Sales Team</p>
```

Text message templates

These templates are used for text (SMS) messages. The maximum number of characters allowed for a text message is 160 characters. A warning message will be displayed if the current message is close to or over this limit. Remember to allow enough room for dynamically inserted information, such as the customer reference.

Example 13.3. Example SMS template

```
Your document (Ref: %%custref) is ready. Please visit
http://mysite.com/login and log in to download your document.
Sales Team
```

Default and queue-specific templates

As well as the set of default templates, you can create and manage templates for a specific queue. If one or more queue-specific templates for a given message type are available for the queue on which a document is processed, those will be selected; otherwise, the default templates will be used.

13.3. Using electronic document delivery

Setting up a queue for electronic document delivery

Electronic document delivery may be enabled for a given queue on the Archive page of the queue configuration section (see “Archiving” in Chapter 6: *Queue Management*).

Go to **Print Platform > Queue Config**, select the queue from the dropdown menu and click on the **Archive** tab. To enable electronic delivery, first make sure that archiving is enabled for the queue. Archiving is necessary for electronic delivery, because documents are archived whilst awaiting delivery. Check the "Deliver electronically" checkbox and click on **Update** to save the settings. Print jobs on the configured queue will now be handled by the electronic delivery system.

Note

The archive feature of DataGateway is used to make a store of the documents for electronic delivery. For this reason, archiving must be enabled on the queue for electronic document delivery to work.

Electronic delivery may be applied on the queue that processes the incoming print data, or it may be deferred and applied on a later queue in the print chain.

Print data for electronic document delivery

Electronic delivery can be used either with structured input data (see “Using structured input data”) or with pre-formatted documents.

When using structured input data, the delivery parameters, such as name, email address and mobile number of the intended recipient, are extracted from specified fields within the data. The keys to be used for these details are specified in the electronic document delivery configuration.

DataGateway's document content control system (see Chapter 9) may be used to extract delivery details from incoming formatted print data, allowing electronic delivery to be applied to documents generated by many different applications. For more details, please contact WML.

Delivery of electronic documents

Documents processed by the electronic document delivery system are stored to a system archive. Information about each document is recorded to a database, which is checked periodically to determine whether there are any pending notification messages.

The customer will receive notification of a new document by email, SMS, or both, depending on the message preference configured. The wording of the notification messages comes from message templates, which are set up in the electronic document delivery user interface (see “Message templates”). At least one message template needs to be set up for each type of notification in use.

Figure 13.2 shows an example notification email. The customer reference and access code in this example have been inserted dynamically from details in the electronic documents database.

Figure 13.2. Customer notification email



Repeat notification messages are sent according to the electronic document delivery configuration, which defines the number of messages to be sent and the number of hours between messages. Different message templates may be set up for repeat notifications, if required.

Email attachment

If the chosen message preference is "Email attachment", then the customer will receive an email message with the document included as an attachment. They will be able directly to view and print the document.

The status of the document in the database will be set to *Emailed as attachment* and delivery will be considered to be complete.

Secure delivery

Message preferences of "Email", "SMS" and "both SMS and email" correspond to the *secure delivery* route. Secure delivery involves the recipient either logging in to an existing customer system or accessing a secure portal on the DataGateway. The stored document is linked to the recipient's account or identification details, and will be made available for download on successful authentication.

Note

See Section 13.4 for more information about setting up the web portal and integrating it into your own website.

The status of the document in the database will reflect the customer's success in downloading the document. Delivery will not be considered to be complete until the document has

been downloaded successfully by the customer or the admin user has printed the document for mail. The system can be set up to revert to hardcopy if a document remains incomplete after a given time period.

Using the customer support interface

The **Electronic Delivery > Search** page provides a customer support interface for the electronic delivery system. This allows customer support staff to search documents that have been processed by the electronic document delivery system, and to check on the delivery status of a given document or to troubleshoot any problems that a customer may be having with gaining access to their document.

A user can search for a particular document using any of the customer delivery details (name, date, reference, email address or mobile number), or by using the access code generated for the secure portal delivery method. The search results show a summary of information for each matching record, and the delivery status of each document (Figure 13.3).

Figure 13.3. Searching the electronic document delivery database



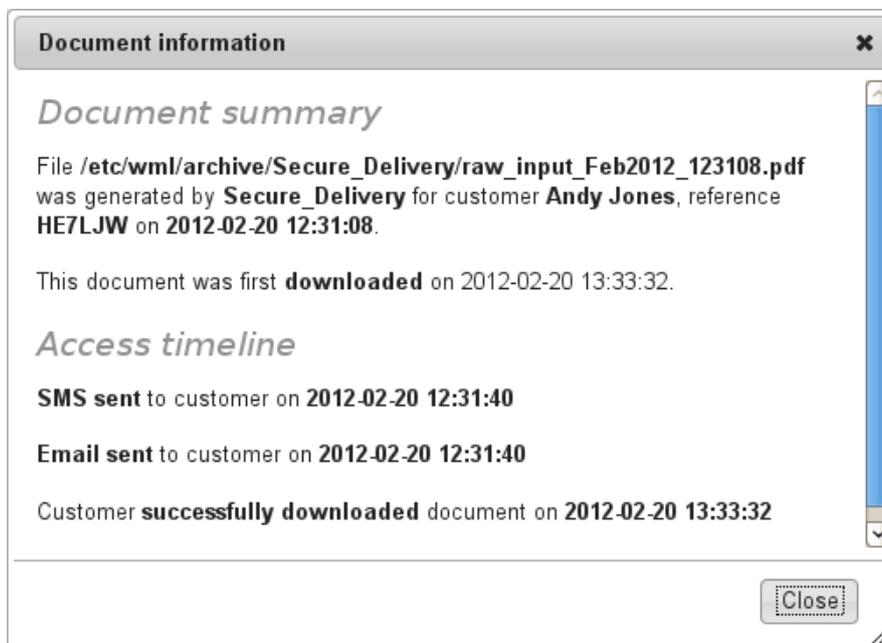
The status of a document may be one of the following:

- Not yet delivered** The document has not yet been downloaded successfully by the customer.
- Downloaded** The customer has downloaded the document successfully.
- Attachment emailed** The document was delivered by email attachment.
- Printed for mail** The document has been printed for delivery by mail.
- Cancelled** Delivery of the document has been cancelled by an admin user.

Unless the status of a document is *Not yet delivered*, the time at which the status changed is given in the status column.

Click on the information icon (ⓘ) next to the status message to bring up a full summary of document information, together with the access timeline for the document (Figure 13.4).

Figure 13.4. Document summary and access timeline



The document summary gives an overview of the document and its current status. The access timeline gives details of notification messages, download attempts and admin actions on the document.

Document operations

Once a document has been located (Figure 13.5), a number of different operations may be performed on it.

Figure 13.5. Viewing an individual entry in the electronic documents database

Name	Creation date	Customer ref	Mobile/Email	Access code	Status	Action
Andy Jones	2012-02-20 12:31:08	HE7LJW	07730868686 ajones@epl.com	36PF	Downloaded at 2012-02-20 13:33:32	   

Resend notification

If required, a notification message can be resent to a customer by clicking on the send icon (). If the message preference is for both SMS and email, then a dialog will be displayed to allow the admin user to choose which type of message to send. Otherwise, the message will be sent immediately.

Note

If the document has been cancelled, notifications cannot be resent.

View document

A document may be viewed by clicking on the view button (📄). The document will be opened for viewing in another window, so pop-ups should be enabled. A direct link to view the document is provided in case the operation to open the new window fails.

The access timeline will be updated to indicate that the document was downloaded by the admin user. This will not change the delivery status of the document.

Print document

A document may be printed, either for internal use or to be mailed to the customer, by clicking on the print button (🖨️). A dialog will be displayed to allow the admin user to choose the type of print action. If **Print for Delivery** is chosen, the document will be marked as *Printed for mail*. **Print to View** leaves the document status as *Not yet delivered*.

Note

To print a document, a print queue must be specified. This is done by setting the **Output queue** parameter in the electronic document delivery configuration.

The access timeline will be updated to indicate that the document was printed, with reference to the selected print action. If the document was printed for mail and delivery of the document was not yet considered to be complete, then the status of the document will be set to *Printed for mail*. This means that delivery of the document is considered to be complete.

Cancel document

A document may be cancelled by clicking on the delete button (🗑️). After a document has been cancelled, the customer will no longer be able to download it.

The status of the document will be updated to indicate that it has been cancelled.

13.4. Web portal for electronic delivery

Setting up the electronic document delivery web portal

If you have configured the DataGateway to provide documents via a secure web portal, then you will be sending the customer an invitation to access a PDF document via their web browser. This web portal allows the document stored on the DataGateway to be accessed by your customer. It can be integrated into your own website, with the DataGateway providing only the backend functionality of serving (and recording the download of) the PDF document.

The components of the web portal system are:

Customer-facing portal

A customer-facing portal, consisting of HTML and javascript only. This is the page which the customer first lands on when they click the link in their email or SMS invitation.

This page will contain a form which your customer should fill with any additional credentials you want them to provide to authenticate their access of the document. Below the form is the "Get Document" button.

The customer-facing portal need not be hosted on the DataGateway system, *but it must reside on the same internet domain*. For example, the customer-facing portal might have the address <https://www.company.com/documents.html>, with the DataGateway being accessible as <https://datagateway.company.com/>

Portal backend: eddportalops.cgi

When the customer presses the "Get Document" button, the javascript on the customer-facing portal makes a background HTTP request (an AJAX call) to a program called **eddportalops.cgi** on the DataGateway. Continuing the example from the previous item, this would be a call to <https://datagateway.company.com/eddportalops.cgi>

The request contains the credentials on the form, along with the document reference (which will have been carried over from the document invitation). If the provided credentials match the document reference, eddportalops.cgi will respond with a link to the PDF. This may, for example, be <https://datagateway.company.com/view/edoc-abcdef-abcdef-abcdef.pdf>.

The customer facing portal will then display this PDF file in the customer's browser. As the customer's browser accesses and then completes the download of the PDF, the DataGateway records these events for the electronic delivery audit trail.

Built-in customer-facing portal

Although you will almost certainly wish to integrate the provided HTML and Javascript into your own website, the DataGateway does provide a plain portal of its own, which is always accessible as **eddportal.cgi**. This can be useful as a reference for building up your own portal.

eddportal.cgi works as a customer-facing portal exactly as described above, making an asynchronous background HTTP request to the eddportalops.cgi to allow the customer to download the PDF document.

Customer portal sample files

You can download the portal sample files from

<http://wmltd.co.uk/eddPortal.zip>

eddPortal.zip contains four files; eddPortal.html, eddPortal.css and eddPortal.js are the components to include in your customer-facing portal. js/jquery.eddportal.js is the core functionality of eddPortal.js, for reference.

eddPortal.html: Customer portal HTML

The HTML for the customer-facing portal is a simple form, contained within a specially named div element called **wmledd_wrapper**:

```
<!-- START WML Electronic Document Delivery Portal HTML -->
<div id="wmledd_wrapper">
```

```
<!-- Displays feedback from the back-end: -->
<div id="EddFeedback"></div>
<div id="wmlredd_custref_field" class="field wmlredd_field">
  <label for="wmlredd_custref">Reference:</label>
  <input id="wmlredd_custref" class="wmlredd_text"
    type="text" value="Enter reference number"/>
</div>

<div class="field wmlredd_field">
  <label for="wmlredd_firstname">First name:</label>
  <input id="wmlredd_firstname" class="wmlredd_text mandatory"
    type="text" value="Enter your first name"/>
</div>

<div class="field wmlredd_field">
  <label for="wmlredd_lastname">Last name:</label>
  <input id="wmlredd_lastname" class="wmlredd_text mandatory"
    type="text" value="Enter your last name"/>
</div>

<div class="field wmlredd_field">
  <label for="wmlredd_code">Code:</label>
  <input id="wmlredd_code" class="wmlredd_text mandatory"
    type="text" value="Enter the code here"/>
</div>

<button id="wmlredd_getdoc" class="wmlredd_button"
  value="Get document" type="button">
  Get document
</button>

</div> <!-- wmlredd_wrapper -->
<!-- END WML Electronic Document Delivery Portal HTML -->
```

There is no special CSS which has to be included for this form, but a couple of suggestions are provided in `eddPortal.css`:

```
/* START WML Electronic Document Delivery Portal CSS */
/*
 * No special CSS required, but you may wish to style
 * the contents of the div with id "EddFeedback" to
 * have a specified border background colour or an icon.
 * "EddFeedback" is used to display status and error
 * messages to the user. E.g. for a dark grey border:
 */
/* #EddFeedback>div { border: 1px solid #8e8e8e; } */

/*
 * Uncomment the following line to hide autocompleted form
 * fields, if desired:
 */
/* .wmlredd_field_completed { display: none; } */
/* END WML Electronic Document Delivery Portal CSS */
```

eddPortal.js: Customer portal javascript

The javascript for the customer-facing portal requires the jquery javascript library (See <http://jquery.org>), jquery-ui (<http://jqueryui.com>), a third party library called jquery.scrollTo.js and two supporting WML javascript libraries - jquery.wmldbg.js and jquery.fwebui.js.

These libraries are listed in eddPortal.js, which can be incorporated into your web site's HTML as it stands, or you can extract the javascript files from eddPortal.js and reference them using your own <script> tags.

If your webpage already includes the jquery library code, you can modify eddPortal.js to remove the duplicated library code.

The actual javascript code for the portal, excluding the supporting libraries (js/jquery.eddportal.js), is quite short:

```
/*
 * Electronic Document Delivery Document Portal
 *
 * Copyright 2011-2013, William Matthew Limited.
 *
 * Authors: Seb James <sjames@wmltd.co.uk>
 *          Tam James <tjames@wmltd.co.uk>
 */

(function($) {

    $.fn.eddPortal = function (options) {

        // Set up mandatory fields. It's important to call this first,
        // before filling in values from URL variables, so that those
        // values register as having completed mandatory fields.
        $("#wmlledd_wrapper").setupFields();

        // If custref specified, hide wmlledd_custref_field. The URL
        // variable is added to the post data on call to viewDocument.
        // N.B. This could be modified to use the same method as the
        // other fields.
        if ($.getUrlVar('cr') === undefined
            && $.getUrlVar('custref') === undefined) {
            // Don't hide
        } else {
            $("#wmlledd_custref_field").hide();
        }

        // For the remaining fields, fill in the value of any fields
        // specified as URL variables.

        // Adds a class 'wmlledd_field_completed' which can be used to hide
        // the fields via CSS, if required.

        if ($.getUrlVar('code') === undefined) {
        } else {
            $("#wmlledd_code").val ($.getUrlVar('code'))
                .parent (".field").addClass ("wmlledd_field_completed");
        }
    }
});
```

```

if ($.getUrlVar('fn') === undefined) {
} else {
    $("#wmlledd_firstname").val ($.getUrlVar('fn'))
    .parent (".field").addClass ("wmlledd_field_completed");
}

if ($.getUrlVar('ln') === undefined) {
} else {
    $("#wmlledd_lastname").val ($.getUrlVar('ln'))
    .parent (".field").addClass ("wmlledd_field_completed");
}

$("#wmlledd_getdoc").click (function() {
    // Build up a request to the backend to obtain a document to view.
    viewDocument();
});

$("#pageContent").keyup (function(e) {
    var key = e.charCode ? e.charCode : e.keyCode ? e.keyCode : 0;
    if (key == 13) { // 'enter' key
        // Instead of $("#wmlledd_getdoc").click(), just:
        viewDocument();
    }
});
$("#pageContent").fwebuiFileDialog();
}

// Customer reference obtained directly from URL.
// First try to get the "cr" variable.
// Fall back to the more human readable "custref".
function viewDocument () {

    if (!$("#wmlledd_wrapper").checkFields()) { return; }

    var custRef = $.getUrlVar('cr');
    if (custRef === undefined) {
        custRef = $.getUrlVar('custref');
    }
    if (custRef === undefined) {
        custRef = $("#wmlledd_custref").val();
    }
    $.log("custRef is '" + custRef + "'");

    // Set this to any additional prefix to add before the returned file URL.
    // Usually, you will proxy from "www.yoursite.com/" + pathPrefix through
    // to the DataGateway. Suggestion: set pathPrefix to "docs/".
    var pathPrefix = "";

    var postData = "action=viewDocument"
        + "&custref=" + custRef
        + "&referrer=" + location.href // referrer is unused in eddportalops.cgi at present.
        + "&firstname=" + $("#wmlledd_firstname").val()
        + "&lastname=" + $("#wmlledd_lastname").val()
        + "&code=" + $("#wmlledd_code").val();
    $.ajax ({
        data: postData,

```

```
async: false,
url: pathPrefix+"eddportalops.cgi?action=viewDocument",
dataType: "xml",
type: "POST",
success: function (response){
    if ($("#fileurl", response).text() != "") {
        // fileurl is relative with the form "view/edoc-aaaaaaaaaaaaaaaaaaaaaaaa.pdf"
        // fwebuiFileDialogShow() does a window.open() on the passed in path, and
        // also creates a fall-back dialog with a link to the document.
        $("##fwebuiFileDialog").fwebuiFileDialogShow (pathPrefix
                                                    + $("#fileurl", response).text());
    }
    if ($("#error", response).text() != "") {
        $("##EddFeedback").fwebuiError ("Failed to get document: "
                                        + $("#error", response).text());
    }
},
error: function (error_message) {
    $("##EddFeedback").fwebuiError ("Failed to get document. Error code: "
                                    + error_message.status);
}
});
})(jQuery);
```

Cross-site scripting issues

All web browsers will prevent a javascript script on a web page from making an HTTP request to a page on a different domain. This is to combat a common technique used to hijack websites.

Because of this limitation, it is essential that the eddportalops.cgi program on the DataGateway is accessible on the same web domain as the customer-facing portal page.

Most web server software can be configured to redirect requests to certain URLs, and you may need to use this redirect functionality to allow the DataGateway to be accessed on your normal website domain. For example, see the **ProxyPass** and **Redirect** configuration parameters for the Apache HTTP server.

Chapter 14. WML Secure Forms

WML Secure Forms is a module for DataGateway which extends the overlay template editor (see “Overlay template manager”). It presents a view of an overlay template canvas which can be filled in by a customer or by a sales assistant working with the customer.

Secure Forms is enabled by entering a suitable licence key in **System > Licences**.

Typically, an overlay template is set up with a PDF form as the background, providing graphical content, and a set of **variable data** elements. In the Secure Forms interface, each variable element is presented as a text-entry field.

When the form has been completed, the sales assistant selects the **Print** option and the document is printed and archived.

Note

Why *Secure Forms*? Because the graphical content of the form is *not held on the client device*.

For example, Imagine a retail store is using Secure Forms with tablet devices. The sales assistant has customers fill their details into the form using the tablet.

If the tablet is stolen, the ability to reproduce the form (which is likely to be a financial document) is not taken with the tablet; it's still locked in the store's server room.

Secure Forms is implemented using web technology. It's simply a web page. Like the rest of the DataGateway it makes extensive use of Javascript, so client devices must conform to the minimum browser requirements.

Figure 14.1. A template ready to be presented as a Secure Form

The screenshot shows a web browser window displaying a PDF form template. The browser's address bar shows 'Page 1' and various toolbars. The form content includes the WML logo, the title 'SALES CONTRACT', and a subtitle 'This is a demonstration document'. The form fields are as follows:

Name:	<input type="text" value="_name_"/>	Ref No.	<input type="text" value="_reference_"/>
Address:	<input type="text" value="_addr1_"/>		
	<input type="text" value="_addr2_"/>		
	<input type="text" value="_addr3_"/>		
	<input type="text" value="Postcode: _postcode_"/>		
Tel:	<input type="text" value="_telephone_"/>		
Mobile:	<input type="text" value="_mobile_"/>		

At the bottom, a table header is visible:

Product code	Qty	Description		
--------------	-----	-------------	--	--

Figure 14.1 shows a sample template in the overlay template editor. It consists of a PDF form background on which numerous variable data elements are placed, with key names such as name, addr1, addr2 and so on.

Figure 14.2. The Secure Form login

The screenshot shows a login page with the following elements:

- Header: **Print Admin Password** (left) and **Login** (right)
- Text: Please enter the password to administer your device.
- Form: **Print Admin password:** followed by a password input field (displayed as dots) and a **Login** button.
- Text: Or go to: [System Login](#)
- Footer: **Powered by WML** (left) and [Terms and Conditions](#) (right)

To access the Secure Forms interface, log out of any DataGateway system admin session you have. Redirect your browser to <http://DataGateway.address/formportal.cgi>. You will see a login page as shown in Figure 14.2. The password is, by default, the same as for the system admin interface (i.e. “blankdoc”).

Figure 14.3. The Secure Form interface

The screenshot shows a web browser window with two tabs: "WML Dynamic Web to Print f..." and "WML - Innovative Technolog...". The page title is "WML Secure Forms" and there is a "Logout" link in the top right. Below the browser window, the form interface is displayed. It features the WML logo on the left and the text "SALES CONTRACT" on the right. A horizontal line separates the header from the main content, which includes the text "This is a demonstration document".

The form contains several input fields:

- Name:** Mr Steve Smith
- Address:** Three stacked empty text input fields.
- Postcode:** One empty text input field.
- Tel:** One empty text input field.
- Mobile:** One empty text input field.
- Ref No.:** One empty text input field.

At the bottom of the form is a table with the following structure:

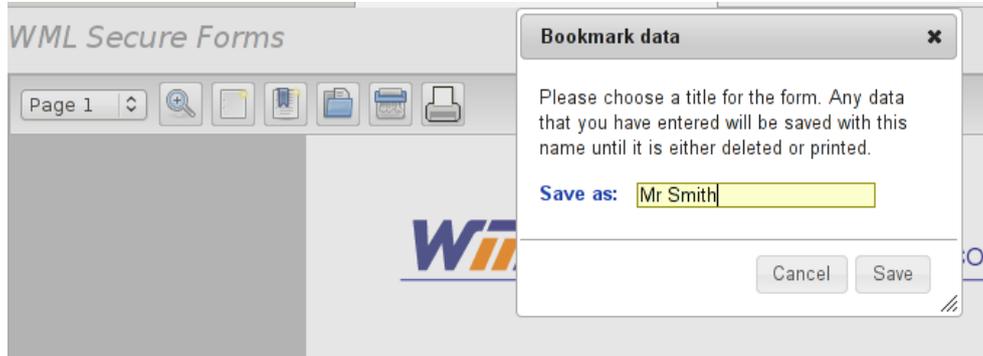
Product code	Qty.	Description	£	p

The browser status bar at the bottom shows "Done" on the left and "wmlvpp.wml" on the right.

Figure 14.3 shows the Secure Forms interface which would be presented to the sales assistance and the customer. It's a very simple interface with the form laid out as it would print and a top toolbar.

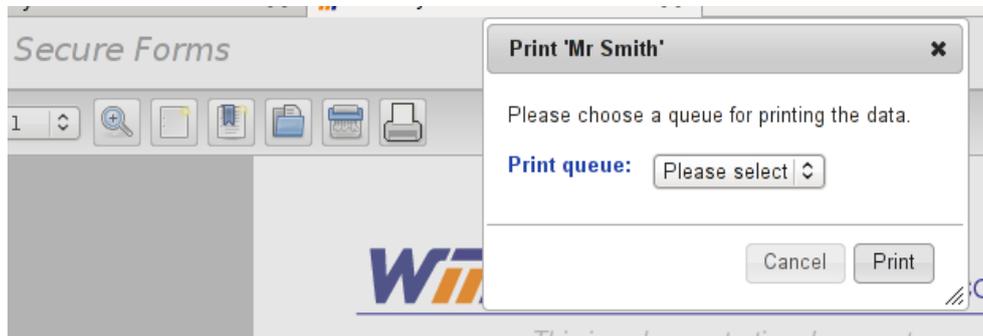
The form is initially presented to fit the page-width of the browser window. As soon as one of the text fields is selected, the form "zooms in". Click outside any text field and the view zooms out again. There is a separate button to action the zoom on the toolbar (🔍).

Every time an element is filled in, the semi-complete form is saved on the DataGateway. The saved form is automatically named "Untitled data", "Untitled data(1)" etc. As soon as the form is printed, the saved form data is erased.

Figure 14.4. Bookmarking an incomplete form

It's possible to name the current form data using the bookmark button (🔖). To load a previously bookmarked form, use the open button (📄).

If a form is incomplete, but is no longer required (i.e. it will not be printed), delete the data using the shred button (🗑️). To start a new named form, press the new button (📄) and provide a name.

Figure 14.5. Printing the form

When the form is complete, press the print button (🖨️). A dialog will open allowing you to select the destination printer (this is a list of all the queues which are configured on the DataGateway). Click **Print** and the form will print to the selected printer.

Note

Most installations of Secure Forms will require customisation to add database interactivity, either to store customer details or to automatically retrieve information for existing customers.

Please contact WML to discuss your individual requirements for a Secure Forms installation.

Chapter 15. Kyocera Prescribe Emulation

The Prescribe emulation is offered as an optional module for DataGateway. The emulation can be enabled by the use of a suitable key in **System > Licences**.

WML's Kyocera Prescribe Emulation converts Prescribe input data, including formatted text and simple PCL, directly into Portable Document Format (PDF).

The converted PDF files can be archived to a document store provided by a Windows or Unix network-share and finally processed for printing through the on-board printer drivers, which convert the document into PCL or PostScript.

The Kyocera Prescribe emulation is configured in the **Queue Config** section of the DataGateway user interface in the tab **Kyo Settings**. Each print queue can be configured with unique Prescribe parameters.

The emulation parameters specified in **Kyo Settings** are effectively the “printer defaults”. The parameters should be configured to match the settings on the printer on which the data was previously being rendered.

The WML Kyocera Prescribe emulation follows the definition of the Prescribe page description language given in the Prescribe Technical Reference 4.51, available from Kyocera.

Figure 15.1. Prescribe emulation configuration interface

Input Text Failover Alerts PPD Driver PPD Options **Kyo Settings**

Prescribe Emulation Settings

Configure the Prescribe Emulation parameters here.

Orientation: Portrait

Duplex mode: Simplex

Top margin (cm): 0.84582

Bottom margin (cm): 27.9984

Left margin (cm): 0

Right margin (cm): 19.794

Paper width (cm): 21

Paper height (cm): 29.7

Bottom clip (cm): 28.9984

Right clip (cm): 19.794

Default resolution: 300dpi

Country code: US ASCII

Default cassette: CASS 1

CASS 1 digital stock: wml_letterhead.pc

CASS 2 digital stock: Specify a tag Letterhead

CASS 3 digital stock: Specify a tag Plain

CASS 4 digital stock: Specify a tag

Character set: ISO 6 ASCII

Persistent state: (Save printer state and macros between jobs)

Update

15.1. Basic Defaults

The basic parameters, *orientation*, *duplex mode*, *default margins*, *paper width* and *height* and *clipping* parameters, will be familiar to all users and function according to the Prescribe specification.

The *country code* and *character set* are used to select the appropriate internal character set and glyphs according to the Prescribe specification, where possible.

15.2. Paper Cassettes

The *default cassette* is the paper cassette which the emulation will use if the cassette is otherwise unspecified (with a CASS or PSRC command).

The DataGateway has two ways to apply paper stock to your printing. It can either apply *digital paper stock* to the output, using a PDF file as the background to the print, or it can embed *PostScript paper selection tags* into the output data.

Note

When used with a PCL printer, only digital letter stock can be used to emulate the CASS/PSRC Prescribe commands.

Using Digital Paper Stock

Digital stock is especially useful when building electronically distributable documents. Digital stock can be used to add a letterhead or the company logo to documents which were previously printed on pre-printed stock.

To set digital stock for a particular cassette, you must first upload the PDF of the stock onto the DataGateway appliance. Got to **File Manager > Form Files**. Use the "Upload a new file" field to upload the PDF. Newly uploaded PDF files will now appear in the menus in **Kyo Settings**. To set your newly uploaded PDF file as the stock for Cassette 1, select it from the menu for "CASS 1 digital stock".

Using PostScript Paper Selection Tags

To have a paper selection tag embedded in PostScript output for Cassette 1, select "Specify a tag" from the menu for "CASS 1 digital stock". Type the exact text of the tag into the adjacent text field.

Note

Common text tags are "Plain" and "Letterhead" but be aware that these are often case sensitive and each printer will have its own tags. In some cases, the tags can be user-specified in the printer or RIP environment.

15.3. Persistent State

If "Persistent state" is checked, then the WML emulation will save the effective printer state between print jobs. This ensures that Prescribe macros are stored between jobs. It will also store the current paper cassette value, page margins and so on.

Storing and re-reading the printer state between print jobs has a small performance cost.

If you need your Prescribe macros to be stored between individually sent print batches, then check "Persistent state", otherwise leave it un-checked.

Chapter 16. Handling output from CDL Strata

16.1. Extracting CSV from Strata PostScript output

This section describes how to consume data generated by the Cheshire Datasystems Limited (CDL) Strata system and use WML's on-board Template Manager to create documents.

Because the CDL Strata system will output documents only in PDF or PostScript format, it is necessary to produce output from Strata containing a text representation of variable data in CSV format. This is essentially a print-out of the CSV.

CSV-formatted text, encoded within PostScript files can be extracted by the filter program **stratapstocsv**.

To enable **stratapstocsv** and print one stream of documents from a CDL Strata system requires two queues on the DataGateway. The input queue accepts PostScript, invokes the **stratapstocsv** filter, which outputs a PDF status or error page, and if successful, it creates a print job of the extracted CSV on the output queue.

The output queue acts as a standard template-creation queue, building a document using the CSV input data and on-board templates.

16.2. Strata input queue

In order to cause PostScript files to pass through the **stratapstocsv** filter, rather than the default **wmlgspstopdf** filter, it's necessary to change a file which controls the conversions between different formats. Go to **Print Platform > Platform settings > mime.convs**. Find the line which reads

```
application/postscript application/pdf 60 wmlgspstopdf
```

Replace this with:

```
application/postscript application/pdf 20 stratapstocsv
```

Now save **mime.convs** then go to **Print Platform > Platform settings > Cups Control**. Press **Restart CUPS**.

This ensures that any incoming PostScript is passed through **stratapstocsv**. The PDF status output of **stratapstocsv** is then passed to either your printer, or into a destination share or the null output. For testing it is helpful to pass the status output to a share in PDF format, so select a PDF PPD in **Queue > Queue Config > PPD** (for your input queue).

The output queue to which jobs should be sent is configured in **Print Platform > Queue Config > Text > Output queue**. You can only set this after you have created your output queue.

There are no other special settings for the Strata input queue.

16.3. Strata output queue

The Strata output queue is a standard queue. The only setting which must be made is the template choice. In **Print Platform > Queue Config > Profiles > Use overlay template** select either a static template, or **Triggered**.

16.4. CSV data

stratapstocsv will extract any text between two special lines “WmlInputDataStart” and “WmlInputDataEnd”. Ensure that these lines of text appear in your PostScript document.

The text-extraction program used (ps2ascii.ps) will concatenate texts which it believes to be in the same paragraph. To ensure it separates out your key-values, write out your text with line spacing of 1.5 or 2.

You can set the parameters for the CSV-parsing in **Print Platform > Queue Config > Text**. This allows you to choose the separator character, enclosure character and page and job separators. For more information about these, see Section 7.7.

Because the CSV text must be laid out in a PostScript document and then extracted in ASCII format from the PostScript, it's important to correctly encode non-ASCII characters using **XML character references**. This is a way to specify the numeric character encoding of a character. Any non-ASCII characters in your keyed values should be presented as an XML character reference.

To determine the XML character reference for a non-ASCII character, first find its Unicode character point. See <http://www.unicode.org/charts/> for the Unicode character points. For example, the Unicode character point for the UK pounds sterling symbol is 00A3 and can be found in the Latin-1 Supplement Unicode chart. To incorporate this character into an XML field it should be written as:

```
&#x00A3;
```

or (the leading zeroes are not necessary):

```
&#xA3;
```

Your key-value pair might look like this:

```
<variable key="amount" value="&#xA3;100">
```

For any Unicode character point, simply prepend “&#x” and append “;” to the Unicode character point.

Don't forget to incorporate a document reference and template in your CSV data using the keys **WmlDocRef** and **WmlTemplate**. The document reference will be used as the job title for the job created on the output queue.

Here's an example CSV file, where the separator character is set to '=':

```
WmlInputDataStart  
WmlBatchRef=Batch1  
WmlDocRef=ABCDE12345
```

Handling output from CDL Strata

```
WmlTemplate=TestTemplate
FirstName=Bob
SecondName=Jones
WmlInputDataEnd
```

In the next example, a page separator is used - it's set to "WmlNewPage". The variables which may appear on any page appear before the first "WmlNewPage". After the first "WmlNewPage", any variables apply to page 1 of the document. Note that WmlTemplate is not set until page 1 of the document. After the next "WmlNewPage", variables apply to page 2 of the document.

```
WmlInputDataStart
WmlBatchRef=Batch1
WmlDocRef=ABCDE12345
FirstName=Bob
SecondName=Jones
Banner=Original
WmlNewPage
WmlTemplate=TestTemplate1
WmlNewPage
WmlTemplate=TestTemplate2
Banner=Customer Copy
WmlInputDataEnd
```

16.5. Errors

The input queue and output queues will generate a PDF describing any error which occurs. Some possible errors include:

<i>Failed to find CSV start or end</i>	The PostScript file was converted to text, but the start of the CSV section ("WmlInputDataStart") and the end ("WmlInputDataEnd") were not found.
<i>The CSV text could not be extracted from the input PostScript</i>	No text could be extracted from the PostScript file. Check for any errors in the PostScript.

Chapter 17. Output Batching

17.1. Output batching introduction

Output batching is one of the “output protocols” available on a DataGateway queue. The jobs processing on a queue with a batched output will be placed in a temporary store, from which they can be released when it is convenient to print the run.

Output batching is intended to help a print production house to manage the production of documents which are “drip-fed” from its customers over a wide-area internet link. The jobs can then be released in one or more batches for efficient fulfilment.

In some cases, several “streams” of documents may be directed into separate output batch queues where it is necessary to handle these separate streams in differing ways. For example, stream “A” may require different pre-printed inserts to be included from those included with the documents printed via stream “B”.

Output batching also allows the DataGateway to be used to implement a “bureau service”, in which transactional key-value data is uploaded to the DataGateway and (after processing with WML Template Manager and Document Content Control) a batch of PDF or print-ready files is downloaded in a zipped archive.

Note

Output batching is available on systems with the production print licence flag (PROPRINT) enabled. Each use of the batch output protocol will consume one system credit.

Production print staff can manually “release” batched jobs to the final output printer using the DataGateway web user interface. Batches can also be automatically released on a scheduled basis or downloaded in a zip archive file.

Automatic, daily release may be suitable for a print operation in which as many documents are produced as possible before a particular cut-off time in the day, after which, print production begins and all those documents which were produced before the cut-off time are guaranteed to enter the mail system that day.

Output batching is configured on a per-queue basis. The per-queue options are configured in **Print Platform > Queue Config** and are described in the section called “Output”.

The resulting documents are reviewed in **Print Platform > Output batching** which provides controls to release or retrieve the print jobs.

The WML Template Manager is used to create output batch templates which can be used to apply additional text to batched documents for production workflow (such as the document sequence number). The Template Manager is also used to create header and trailer templates to print before and after the jobs in the output batch, if required.

17.2. Queue configuration for output batching

Figure 17.1. Output batching: Queue configuration

The screenshot shows the CUPS Queue Configuration web interface. At the top, there is a 'Switch to queue:' dropdown menu set to '01INPUT' and a 'Print status page' link. Below this is a navigation bar with tabs for 'Input', 'Text', 'Profiles', 'Archive', 'Failover', 'Alerts', 'PPD', 'PPD Options', 'Output', and 'Info'. The 'Output' tab is selected.

The main content area is divided into two sections: 'Queue Output Configuration' and 'Production Print Settings'.

Queue Output Configuration: This section contains 'Output Options' with the following settings:

- Queue output protocol: batch (dropdown)
- IPP Compression:
- Queue output address: ramdisk (text input)
- Port/Queue: (text input)
- Diagnosis:
- Suppress error pages:

Production Print Settings: This section contains several sub-sections:

- General:** Page scaling: 100% (slider)
- Imposition:** Apply imposition: None (dropdown)
- Job Ticket/Job Header:** Enable?:
- Output Batching:**
 - Enable?:
 - Output queue: No output queue (download batches) (dropdown)
 - Batch the output: in backend (after PPD) (dropdown)
 - Concatenate jobs:
 - Temp files: RAM only (dropdown)
 - Daily release time: Manual/Threshold (dropdown)
 - Release threshold: 0 (text input)
 - Document template: None (dropdown)
 - Batch header: None (dropdown)
 - Batch trailer: None (dropdown)

An 'Update' button is located at the bottom left of the configuration area.

Configure output batching features for a queue in **Print Platform > Queue Config**, as shown in Figure 17.1.

Enable output batching by checking the first checkbox and choose an output queue. It's necessary to choose the output queue if jobs will be released directly to the printer rip. If no queue is chosen, it will only be possible to download a zipped archive of the output jobs.

Note

To manage re-prints of spoiled documents, turn archiving on in the output queue, and set a suitable expiry on the files. Staff can search for the spoiled document in the archive and produce a direct re-print.

The following options are available to control how output batching is applied in each queue:

<i>Enable</i>	Whether or not output batching is enabled for the queue.
<i>Output queue</i>	.
<i>Batch the output</i>	This is a menu to choose whether the batching occurs before the output of the queue, or after the output, in the “backend”. If Queue output protocol in Output Options is set to batch , then Batch the output will be forced to in backend (after PPD) . It is possible to enable output batching whilst simultaneously sending documents to another output protocol (such as IPPS, or LPD). To do this, choose PDF Files (before backend) .
<i>Concatenate jobs</i>	Whether or not individual pending jobs are concatenated together into a single file when released. This will concatenate the files regardless of type, so care should be taken that the files which are stored in the output batch are in a suitable format for concatenation (PCL and PostScript are suitable, PDF is not).
<i>Temp files</i>	Determines whether the batched jobs are stored on the RAMdisk or within local storage before release.
<i>Daily release time</i>	If set to Manual/Threshold , then batched jobs will be released manually or when a threshold number of jobs are ready for release. Otherwise, this specifies the time in the day when the system will automatically release pending jobs.
<i>Release threshold</i>	How many jobs must be present before and output batch is released. If set to 0, then any number of files may be batched and no automatic release will occur.
<i>Document template</i>	This is a menu of all the templates stored in the Template Manager. If a template is chosen here, then it will be applied to each job included in an output batch. If set to None , documents are unmodified as they are included in a batch.
<i>Batch header</i>	The template to use for the output batch header document, output before the start of the batch.
<i>Batch trailer</i>	The template to use for the output batch trailer document, output at the end of the batch.

Figure 17.2. Output batching: Header and trailer templates

The template specified in **Document template** is applied to all the pages in the output documents, and allows the document sequence number, the page number, page count, date and time to be applied to each page in the output documents. This may be important to provide a reference from which spoiled documents can be identified for re-print.

To further aid the production process, a batch header and trailer document can be created, containing the number of documents, the sequence start and end numbers and a number of other pieces of information (Figure 17.2). To define how the header and trailer will appear, templates should be set up in the Template Manager and selected in **Batch header** and **Batch trailer**.

Warning

Currently, it's possible to create the header, trailer and modify the documents in the output batch only if they are batched in PDF format.

The process of creating a batch header or trailer template is described later in this section.

17.3. Output batch list

Figure 17.3. Output batching: List of pending jobs

Queue	Output Queue	Applied in	Release Parameters	Output batches pending	Action
01INPUT	No output queue (download batches)	backend	Manual/Threshold	4	
02DCC	No output queue (download batches)	backend	Manual/Threshold	20	
COURIER	FINAL_COUR	backend	Manual/Threshold	1	
ROYALMAIL	FINAL_RM	backend	Manual/Threshold	8	

Enable Cancel

Figure 17.3 shows the **Output Batches** page within the **Print Platform** section of the Data-Gateway user interface. This is the interface which print production staff will become most familiar with.

Warning

The output batching feature has an internal locking system, which will delay the loading of the **Output Batches** page if documents are being processed. If the page seems to be taking some time to load, please be patient and let the operation complete.

On this example system, four queues have output batching enabled. The table lists the output batching queue itself, the output queue, where in the filter chain the output batching was applied, the scheduled release time, the number of jobs pending and some action buttons. The first two queues (01INPUT and 02DCC) have no output queue set, and so they have only the “open” action, which packs the pending jobs into a zip file and allows the user to download that file. COURIER and ROYALMAIL each have an output queue set, so they also have the “release” action available.

Note

If the batched jobs need to be discarded, **Enable Cancel** can be clicked, which will reveal cancel actions for each queue.

Releasing output batches

To select jobs for release in an output batch, select the release action (Figure 17.4) for the selected queue.

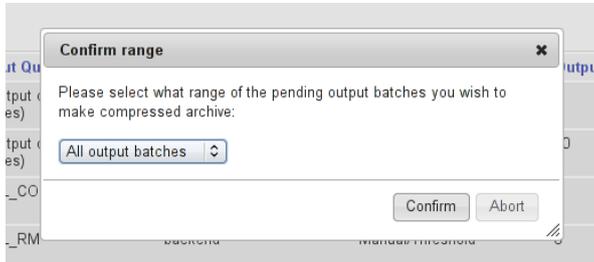
Figure 17.4. Output batching: The release action



When the release button is pressed, a dialog is shown to choose the range of jobs to include in the output batch (Figure 17.5). The options are:

- All output batches* Include all pending print jobs in the output batch.
- Up to cut-off time* Choose a day and time. All pending jobs which were received before this time will be included in the output batch.
- Up to cut-off number* Enter a number, N. The first N pending jobs will be included in the output batch.

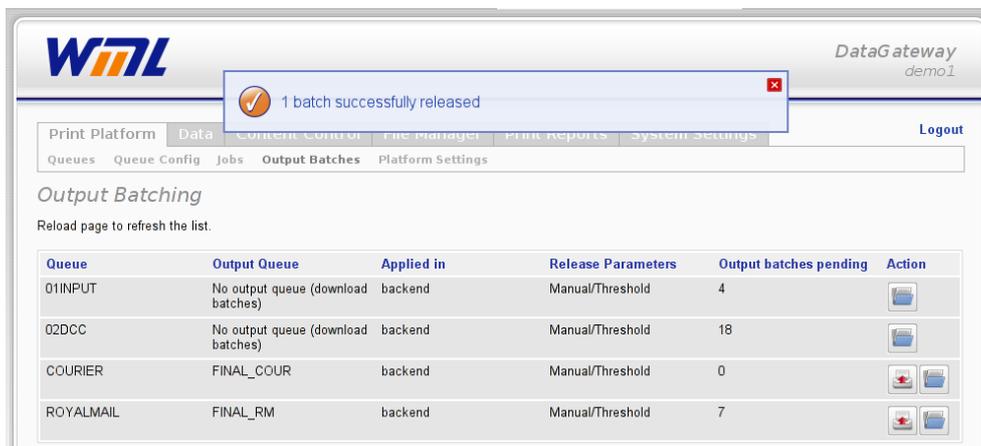
Figure 17.5. Output batching: Confirming the jobs to release



To cancel the release, press **Abort**. The dialog will close and the pending jobs will remain in the output batch list.

To proceed, press **Confirm**. The pending jobs in the range will be spooled to the output queue and removed from the output batch list. A confirmation message will display how many jobs were released to print (Figure 17.6).

Figure 17.6. Output batching: Jobs released to print



Warning

The output batch store is not an archive. Once jobs have been released to the printer, to a rip or downloaded in a zipfile, they cannot be re-released.

To keep a copy of the jobs released in an output batch, turn archiving on in the final output queue.

Downloading output batches

To download pending jobs in an output batch zip file, press the blue folder “open” icon (Figure 17.7). The confirm range dialog will open, as when releasing jobs to an output queue.

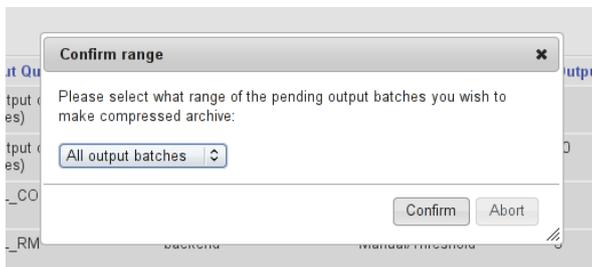
Figure 17.7. Output batching: The download action

Queue	Output Queue	Applied in	Release Parameters	Output batches pending	Action
01INPUT	No output queue (download batches)	backend	Manual/Threshold	4	
02DCC	No output queue (download batches)	backend	Manual/Threshold	20	
03UIP	FINAL_QUEUE	backend	Manual/Threshold	1	

When the release button is pressed, a dialog is shown to choose the range of jobs to include in the output batch zip file (Figure 17.8). The options are:

- All output batches* Include all pending print jobs in the zip file.
- Up to cut-off time* Choose a day and time. All pending jobs which were received before this time will be included in the output batch zip file.
- Up to cut-off number* Enter a number, N. The first N pending jobs will be included in the output batch zip file.

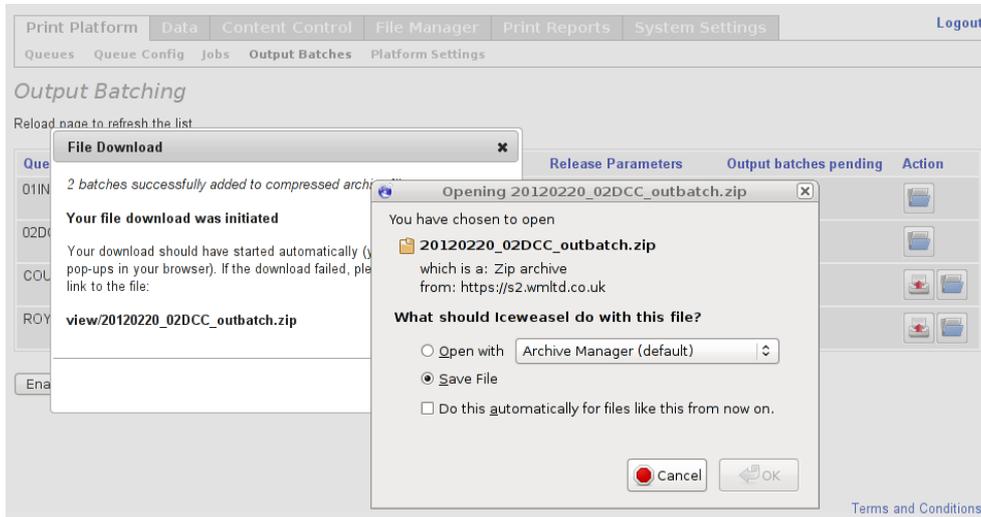
Figure 17.8. Output batching: Confirming the jobs to include in the output batch zip file



To cancel, press **Abort**. The dialog will close and the pending jobs will remain in the output batch list.

To proceed, press **Confirm**. The pending jobs in the range will be packed into a zip file, removed from the output batch list and presented as a download (Figure 17.9) using your browser's own method for downloading and saving files. Often, the option to either save the file or to open it with an application is given, as in Figure 17.9.

Figure 17.9. Output batching: Downloading the output batch zip file

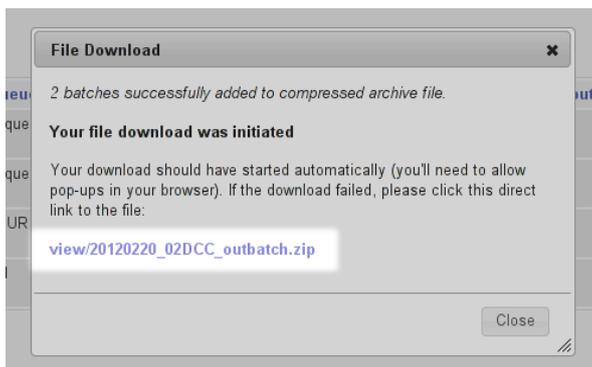


Warning

If you choose to “open with” or “view” the zip file instead of immediately saving it, be sure to save a copy before closing it.

Some browsers will automatically present a download dialog with the option to open or save the zip file. Most browsers will first require the user to “allow popups”. Some browsers (including Microsoft Internet Explorer) make it impossible for a web site to present a download. If you are using one of these browsers, you can simply click the direct link in the WML **File Download** dialog (Figure 17.10).

Figure 17.10. Output batching: Using the direct link to a zip file



If the transfer of the file is interrupted, the direct link can be re-used to make another attempt at downloading the batch file.

If the **File Download** dialog is closed, and batch file has been lost, then you have one more “get out of jail free” card left. *Without re-loading the **Output Batches** page*, click on the blue “open” folder and follow the steps to access the files (press **Confirm** to proceed). The dialog will offer the option of re-downloading the most recent output batch file, as an alternative to creating a new output batch.

Output batch zip file contents

The contents of the output batch zip file depend on the format of the pending jobs, whether a header and trailer are enabled and whether job concatenation is switched on.

Note

Concatenation can only be used with print output formats which can support it. PCL files can be concatenated, as can PostScript, although to concatenate PostScript, it is necessary for the printer to support “job server mode”.

Do *not* attempt to concatenate PDF files. The DataGateway will comply and it will “glue” the files together, but the files will not print or render correctly.

If the queue PPD specifies PostScript output, then the files in the output batch zip file will be given the suffix .ps. If PDF files are produced, the suffix will be .pdf. In all other cases, the files will be given the suffix .prn.

Figure 17.11. Output batching: Contents of an output batch zip file

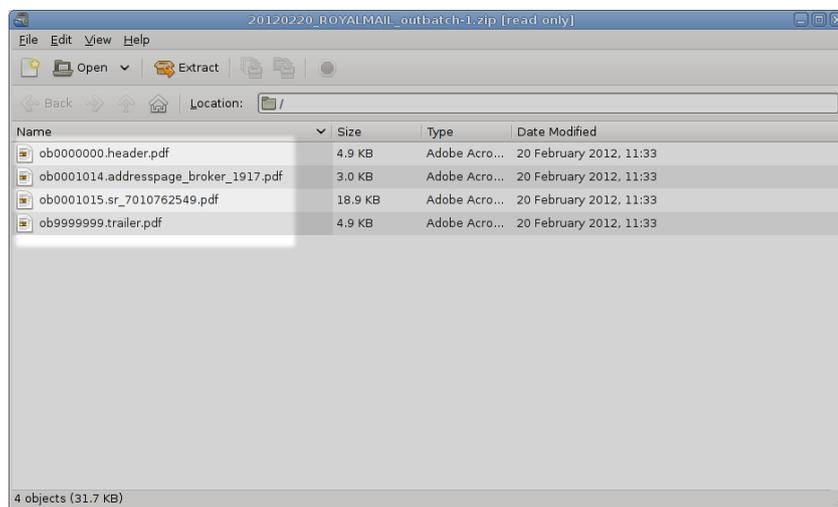


Figure 17.11 shows the contents of an output batch file containing a header, a trailer and two output documents. The header file name is always ob0000000.header.pdf. The “ob0000000” places it first in the list of files. Similarly, the trailer, if present, is always named ob9999999.trailer.pdf ensuring it is last in the list.

All the other files have an “ob” sequence number which gives the order in which the jobs should be output (“ob” for Output Batch). The “ob” sequence number (“NNNNNNNN” in “obNNNNNNNN”) is followed by the job title, then the suffix.

If concatenation is turned on, the output batch zip file will contain a sub-folder “tmp” containing a single file with a name such as “outbatchcat-XXXX.prn” where XXXX is a random string of (hex; 0-9 and a-f) characters such as “2def”.

17.4. Output batch templates

Header and trailer templates

It is possible to specify a user-generated template to use as the header in the output batch. Before selecting a header template, it's first necessary to create a suitable template in the WML Template Manager.

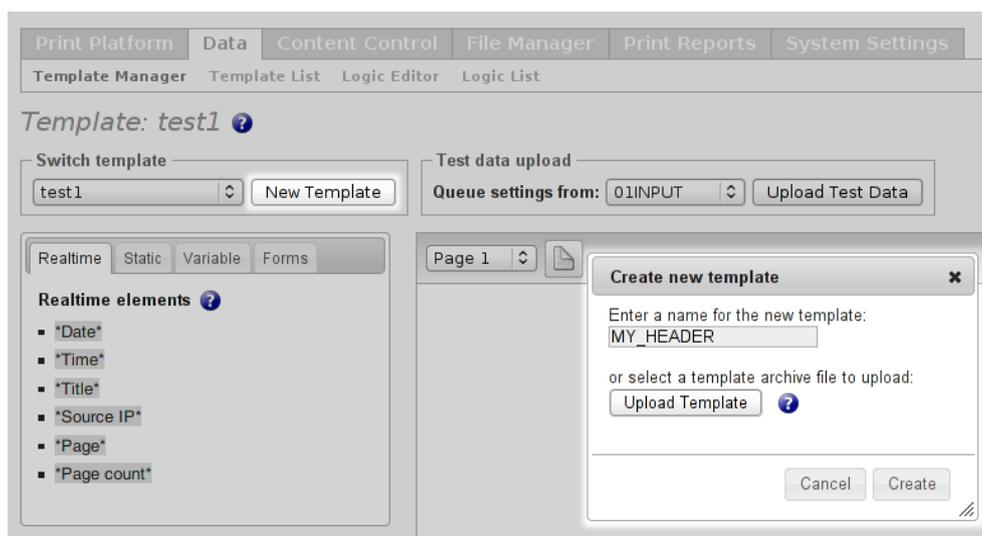
The method described here for creating a header template can be followed to create a trailer template. Just give the trailer a different name, and lay it out according to your requirements.

Note

For a full description of the Template Manager see Chapter 7.

To create a new template, go to the Template Manager at **Data > Template Manager**. Click **New Template** then fill in the name for your header template (Figure 17.12).

Figure 17.12. Output batching: Creating a new template to use as an output batch header



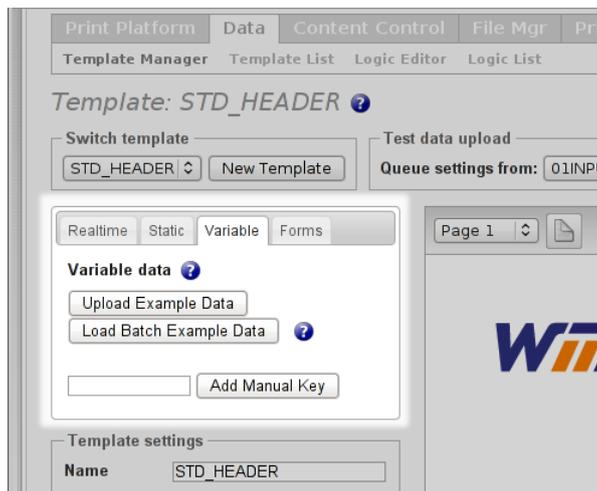
Upload any PDF forms which you will use as the background to your header and enable them in the usual way. You're now ready to lay out realtime, static and variable elements on your template.

Realtime elements are filled in at document creation time using current values. The **Date** and **Time** elements are useful when constructing output batch headers as they will contain the time at which the pending jobs were released in an output batch.

Static elements can be used to add simple text to your template, to augment any text which already exists on your PDF form background.

The output batching system uses a pre-defined set of variable key-value elements to lay out the information about the output batch. It's possible to load an example set of these elements in the **Variable** tab in the Template Manager (Figure 17.13). The button **Load Batch Example Data** will load a list of examples which can be laid out on the current template.

Figure 17.13. Output batching: Loading the example variable data elements



The pre-defined output batch elements are:

- | | |
|-------------------------------|--|
| <i>OutputBatchDocSeqNum</i> | This will show the “ob” sequence number of the current document in the batch. |
| <i>OutputBatchStartSeqNum</i> | This will show the first sequence number in the batch. |
| <i>OutputBatchEndSeqNum</i> | This will show the last sequence number in the batch. |
| <i>OutputBatchPageCount</i> | This will be filled with the total number of pages in the output batch. |
| <i>OutputBatchSize</i> | This will show the total number of documents printed in the batch. |
| <i>OutputBatchQueueName</i> | This element will be filled with the name of the queue to which the output batch was sent. |

Pressing **Load Batch Example Data** will load the list shown in Figure 17.14.

Figure 17.14. Output batching: The example pre-defined output batch keys

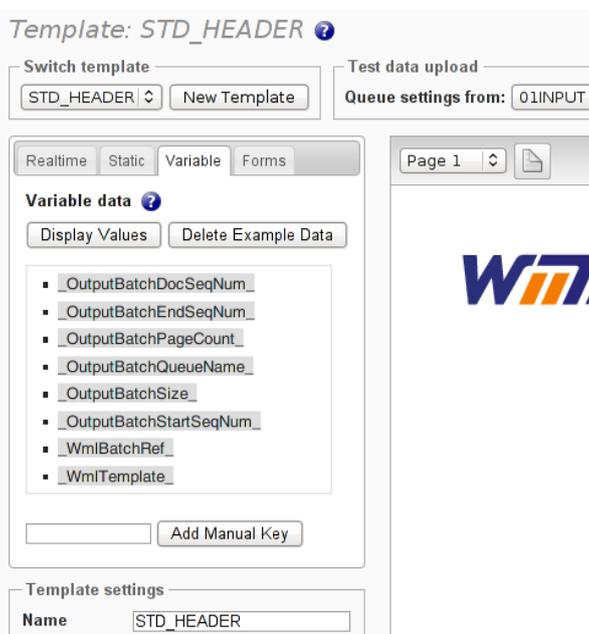


Figure 17.15 shows an example template with each element laid out.

Figure 17.15. Output batching: Editing an output batch header

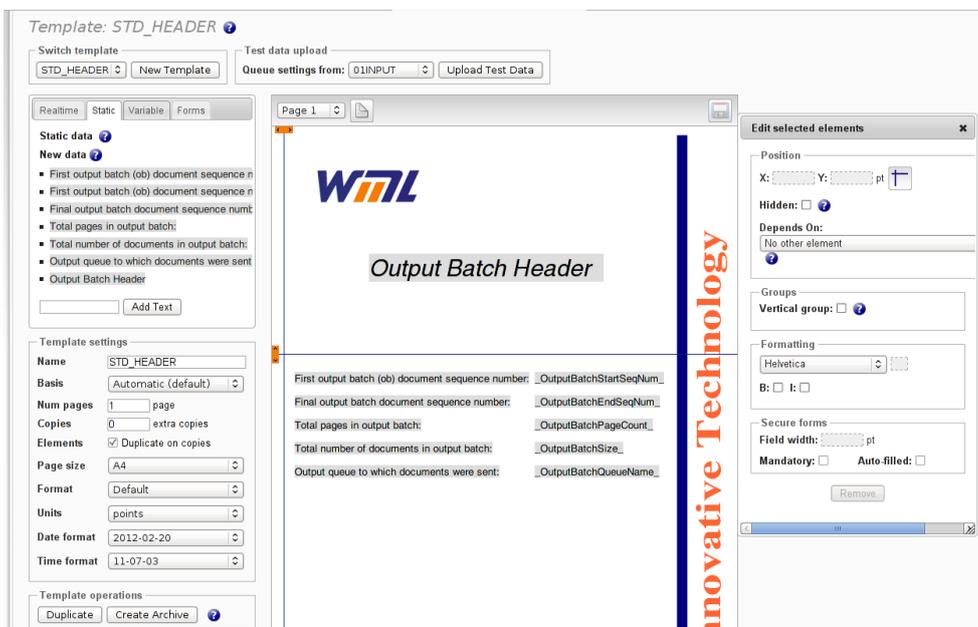
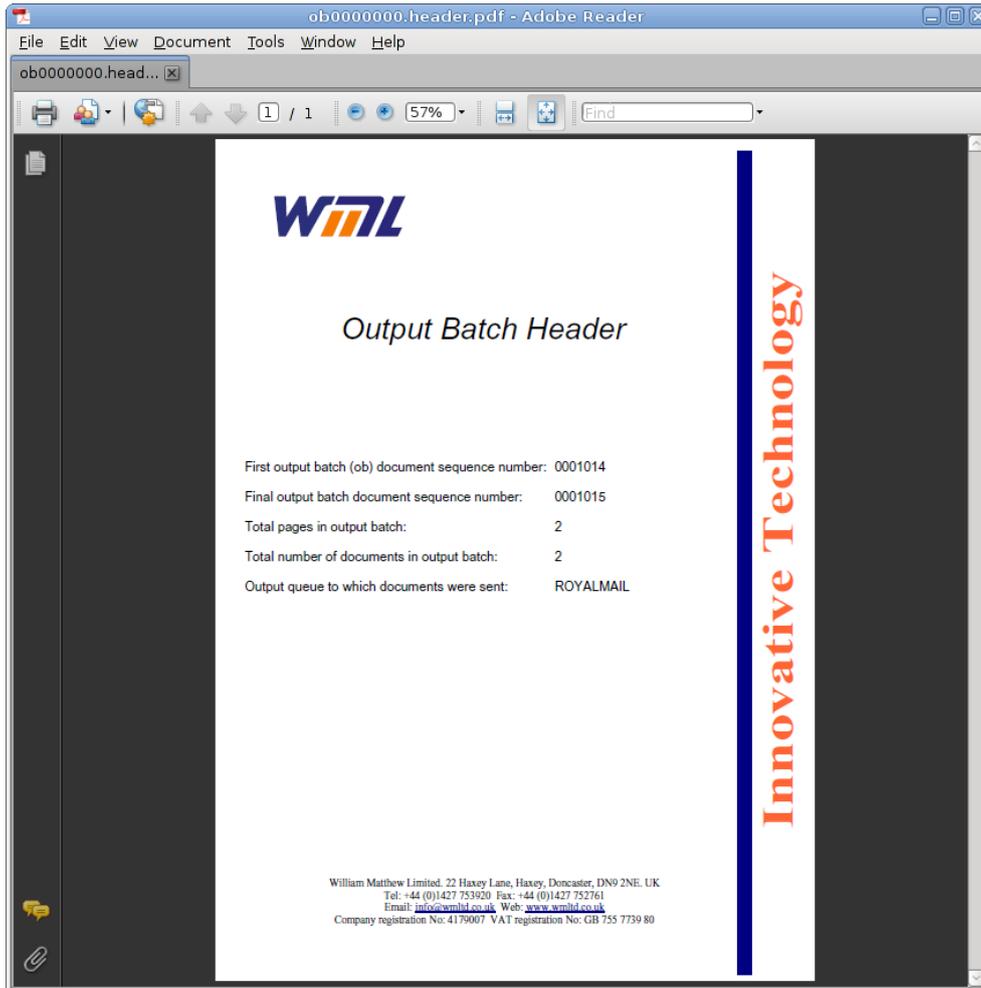


Figure 17.16 shows the resulting header page as viewed in Adobe Reader. This example was printed with a short output batch containing two, single page documents, sent to an output queue called “ROYALMAIL”.

Figure 17.16. Output batching: Viewing an output batch header



Document template

The header and trailer templates are generated from information gathered about the documents in the output batch, but make no changes to the documents themselves.

In contrast, the output batch **Document template** (see Figure 17.2) provides a way to overlay the document “ob” sequence number, page number, page count, date and time on each document when it is released for printing (or download).

The same set of pre-defined keys are available as for building header and trailer templates. See Figure 17.14 for details.

Figure 17.17. Output batching: Editing an output batch document template

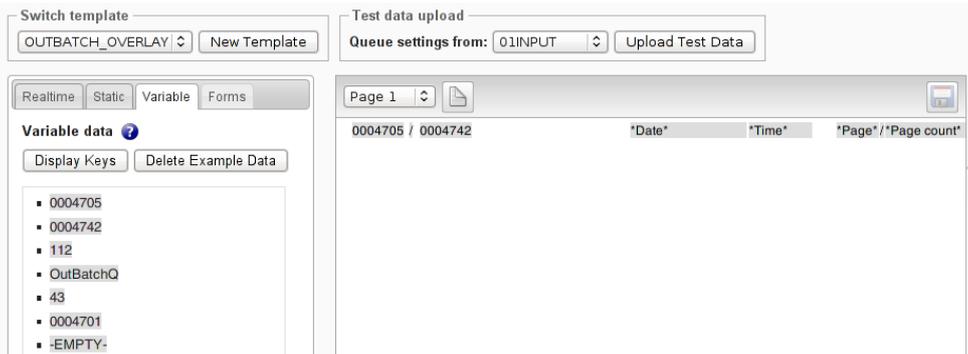


Figure 17.17 shows an output batch document template being edited. It shows realtime elements **Date**, **Time**, **Page** and **Page count**, along with the batch sequence number (**OutputBatchDocSeqNum**) and the final sequence number (**OutputBatchEndSeqNum**). The user has pressed **Display Values** and so the example values for these fields (0004705 and 0004742) are being shown, rather than the key names.

It's possible to specify that elements should appear on an individual page only, or on all pages using the Template Manager's pages menu. In Figure 17.17, the elements will only be added to page 1 of each documents. To have elements added to all pages of each document, select **All pages** in the page menu (see Figure 17.18).

Figure 17.18. Output batching: Selecting all pages so that elements will appear on every page of a document

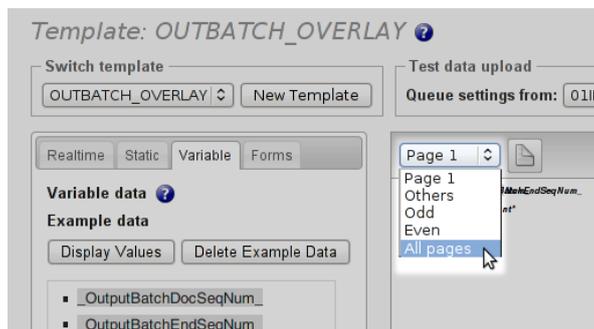
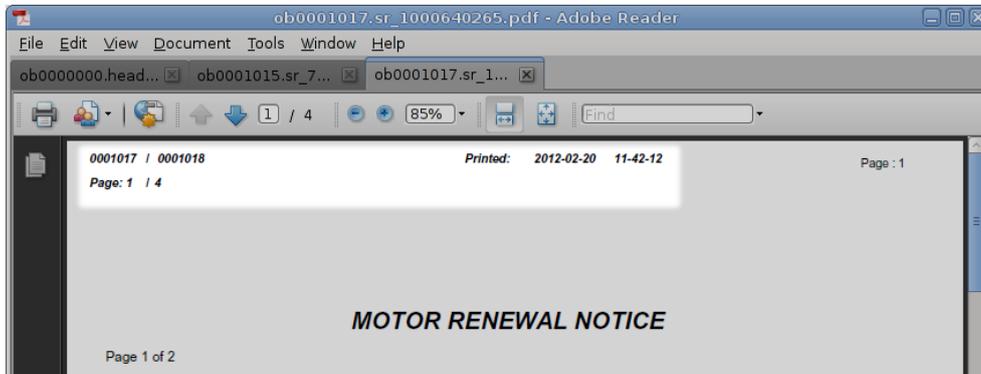


Figure 17.19 shows the result of applying a similar document template to the documents in an output batch. The highlighted region shows the overlaid text.

Figure 17.19. Output batching: Realtime data overlaid on a document at output batch release time.



17.5. Re-printing spoiled documents

The work-flow for re-printing damaged documents will depend on where documents are archived. Here, it is assumed that documents are archived in the final output queue and that the print production house takes responsibility for creating re-prints of spoiled documents. It is possible that documents are archived on a DataGateway installed at the customer site, in which case the customer would access the document from their archive and initiate a re-print following a request from the production house.

Setting up archiving on the final queue

Figure 17.20 shows the **Archive** tab in the queue configuration tab for the queue "FINAL_RM". In this example "FINAL_RM" is the queue to which the output batches will be released, and it sends the documents on to the printer rip as they are produced.

Enable archiving, and choose a **Share Type**. Here, the local DataGateway storage is used and an expiry time of 7 days is set. In this case it will not be possible to make re-prints of documents older than a week.

Note

Ensure that the **Job title** is selected in the **Archive File Name** so that the "ob" sequence number will be searchable in the archived files.

Figure 17.20. Output batching: Set archiving for the final output queue, with a suitable archive expiry time.

Print Platform | **Data** | Content Control | File Mgr | Print Reports | Sys Settings | Logout

Queues | **Queue Config** | Jobs | Output Batches | Platform Settings

Queue configuration for FINAL_RM

Switch to queue: FINAL_RM | Print status page

Input | Text | Profiles | **Archive** | Failover | Alerts | PPD | PPD Options | Output | Info

Archiving

Archive jobs to network attached storage

Share Type: Local Drive

Available SMB/CIFS shared folders: Refresh

Manual path entry:

You may need to enter your username/password in the fields below to discover the shares on your network.

Share username:

Share password:

Archive expiry: 7 days

Update

Archive File Name

The archive file name will be formatted as follows. Use the controls below to customize the format.

lpjob_Feb2112_113431.pdf

Fields

Job title
Date: Feb2112
Time: 113431

Add a field
Date: 2012-02-21

Field separator: Underscore

Suffix: .pdf

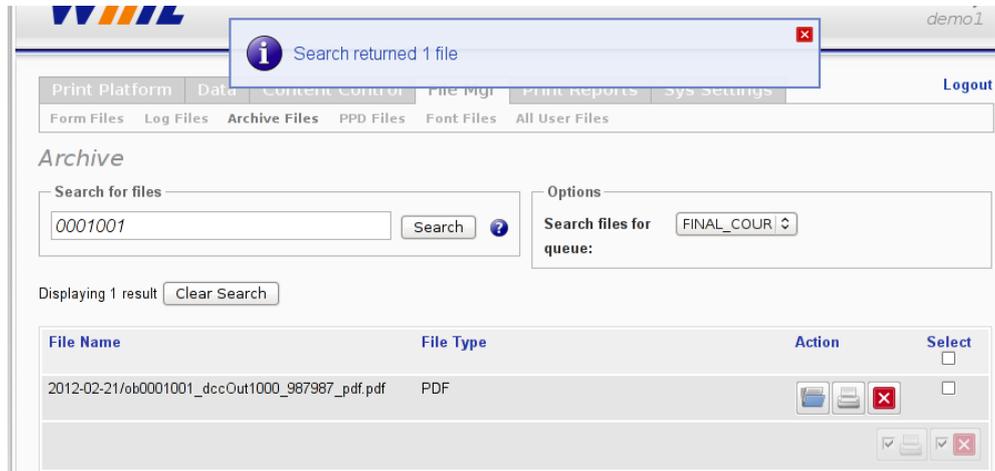
Re-printing the document

To re-print a document, the archive is accessed via **File Manager > Archive Files**. This page shows a simple search box (Figure 17.21). In this example, the search has returned a single document “ob0001001_dccOut1000_987987_pdf.pdf”.

Note

The information following “obNNNNNNN_” is the original title of the document. In some systems, this contains a unique document reference which is viewable on the document. This can be used as the “re-print reference” in situations where the “ob” sequence number cannot be printed on each document.

Figure 17.21. Output batching: Searching for a spoiled document using the “ob” sequence number “0001001”.



To view the document, and verify that it is the correct one, press the blue folder “open” icon; the document should open in a pop-up window or in another tab. To spool it for printing, press the printer icon. A dialog will open in which it is possible to select the queue to which the re-print should be sent.

Note

It is possible to set a security flag in the system to remove the queue menu from this dialog. In that case, the document can only be re-print to the queue on which it would normally be output, or to another fixed “re-print” queue.

Figure 17.22. Output batching: Re-spooling a spoiled document.



Chapter 18. Print Outsource: Sender Site

18.1. Print Outsource Introduction

This chapter describes how to configure your network and firewalls to enable two DataGateway devices to transfer data from one site to another. This technique is typically used to transfer print data generated at one office to a digital fulfilment facility at a separate location - outsourcing the print production.

This chapter will focus on the configuration necessary at the sender site at which print data is generated for production else where. The configuration of the DataGateway sited at the production facility is covered in a separate chapter.

Note

The networking of the DataGateway is very flexible. Configurations described in this chapter can be modified to suit the networking requirements of the site. For example, while many customers use a NAT firewall/router as connection to the public internet, a direct connection without NAT is also possible.

At the sender site, the WML DataGateway is connected both to an internal network from which print data originates and to an external network, which is connected to the public internet. The internal network and external networks may be connected to separate network interfaces on the DataGateway though this is not always necessary.

Some sites will prefer to install a separate, dedicated network connection to carry print data, other sites will prefer to route to the public internet via their existing lines. A standard DSL line is capable of carrying significant volumes of print data, so in most cases an expensive dedicated line is not required.

18.2. Deployment

Figure 18.1 shows the devices which together form the print outsource link. There is a sender DataGateway, a sending firewall/router, a receiving firewall/router and a receiver DataGateway

Note that the DataGateway is always deployed as a virtual print appliance. Each WML DataGateway is a virtual machine image running on a WML Virtualized platform.

This scheme ensures the safety of remote upgrades to the DataGateway firmware. It also allows a single site to host several DataGateway systems on a single physical server.

18.3. External Network

A number of ports must be opened on the source and destination firewalls to enable:

- Data to transfer from source to destination
- Log messages to transfer from source to destination
- Log messages to transfer from destination to source
- SSH data transfer from source to destination
- SSH data transfer from destination to source
- SSH access by support staff to DataGateway virtual print appliance and also to the Virtualized platform.
- Web UI access by support staff

Optionally, the system data on the DataGateway may be maintained using Network Time Protocol (NTP), this may use an NTP server on the public internet. Alternatively a local NTP server may be used.

Figure 18.1. The communications required by the sender firewall

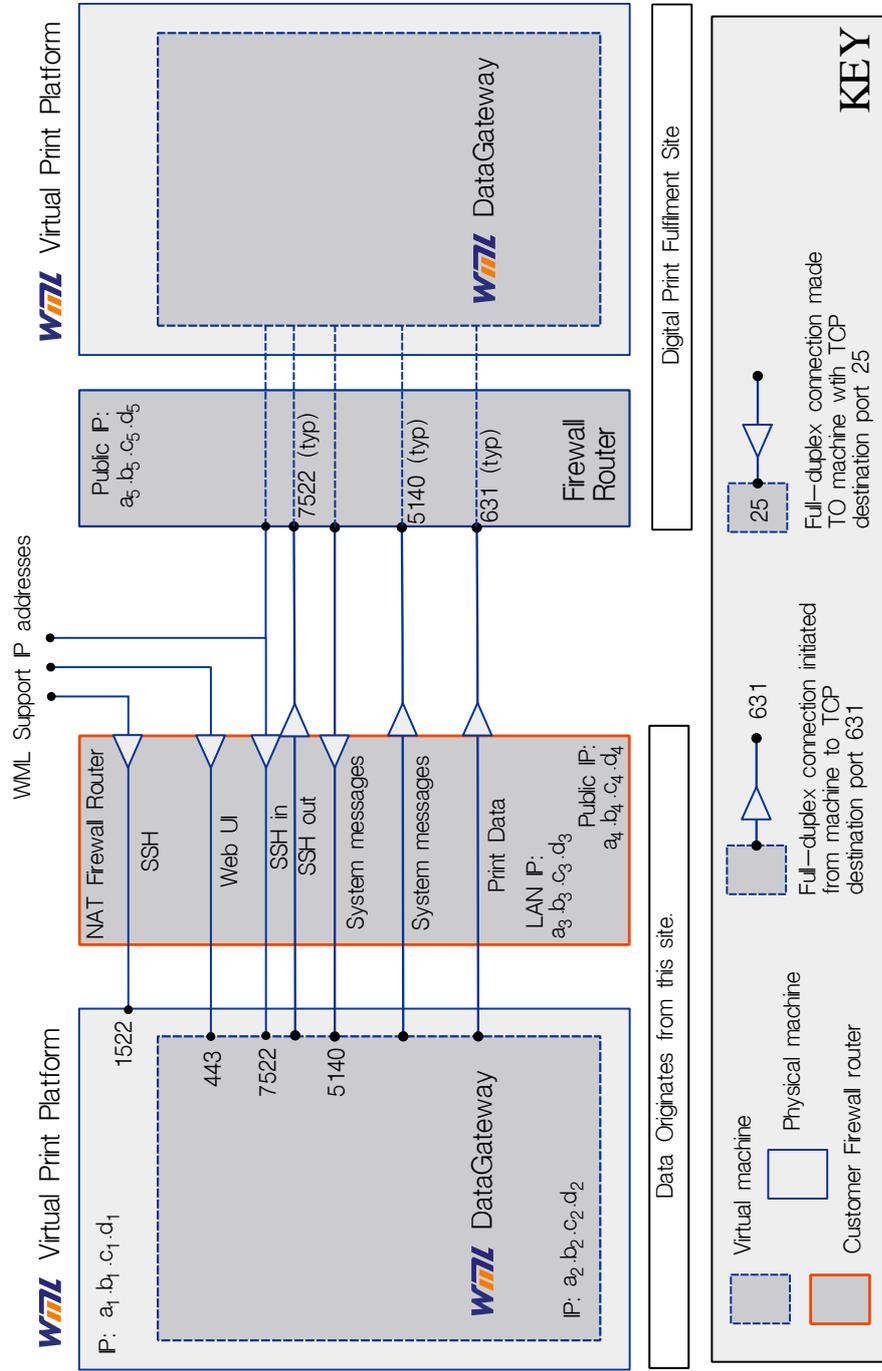


Figure 18.1 shows the traffic which must be allowed to pass through the firewall router at the sender site. The diagram shows connections between the two WML DataGateway devices and connections made to the sender DataGateway from WML support sites.

IP addresses of WML support sites will be advised upon before installation of the system. There are five other IP addresses to note here. a1.b1.c1.d1, a2.b2.c2.d2 and a3.b3.c3.d3 are three addresses on the same IP subnet - the “inside” subnet of the firewall router. a4.b4.c4.d4 is the public IP address of the firewall router - the address to which WML support staff will connect and the address used to send system messages from the receiver DataGateway. a5.b5.c5.d5 is the address of the receiver firewall. The sender firewall must allow outbound connections to a5.b5.c5.d5 to send print data and system messages and allow inbound connections from this IP address for system messages.

Each connection shown on the diagram shows the sending or destination port. If the destination port is given, then the firewall must allow new connections to that destination port on the destination IP address. Return traffic associated with the newly created connection must also be allowed - that is, each connection shown on the diagram is implicitly full-duplex, with data flowing in both directions. The return traffic will be sent to the source port of the created connection. If port 34526 were the source port used to create a connection to destination port 631, then established data returned from port 631 to port 34526 should be allowed to pass through the firewall.

Note

Some ports are variable and may differ from site to site. For example, the destination port on which print data is sent to the receiver is marked “631 (typ)” which indicates that 631 is a typical value for this port. For this traffic, ports 632, 633, 634, etc. are also used.

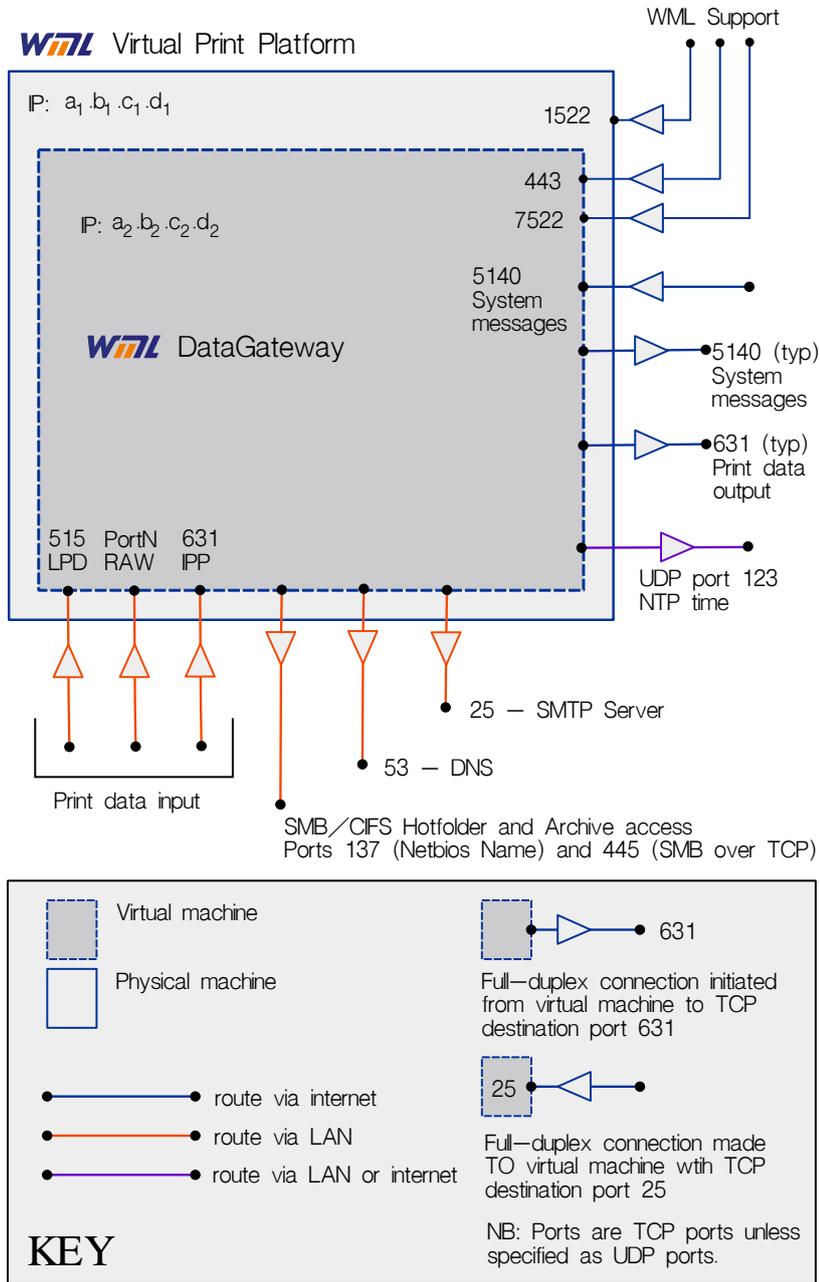
If the source port is shown, then the firewall must allow new in-bound connections to the given port from the relevant IP address or address range.

For example, the firewall router should allow new connections made from a2.b2.c2.d2 to a5.b5.c5.d5 on TCP port 631, as well as associated return traffic.

18.4. Internal Network

Figure 18.2 shows the network connections made to the sender DataGateway. It shows both the connections made to and from the public internet and also the LAN connections.

Figure 18.2. Ports used by the sender DataGateway



Traffic to and from the WML DataGateway on the LAN includes incoming print data, DNS and Netbios naming requests, email messages sent via an SMTP server and SMB (Windows network share) traffic. Network time protocol traffic may optionally be carried on the LAN.

Incoming print data may be transferred using the Berkeley LPD protocol, as a raw TCP data stream (with no fixed port - 9100 is often used) or with the Internet Printing Protocol (IPP). Additionally, an input hot folder may be used. In this case, the DataGateway makes a connection OUT to a SMB/CIFS (Windows or Samba) server and scans a shared folder for print data files, which it then collects and prints.

If necessary, the DataGateway can make use of a DNS server on the LAN. If email alerts or reporting are required, an SMTP server should be made available to the DataGateway on the LAN.

Note

Although it is possible to use an SMTP server on the public internet to send email alerts, this is not advisable. One common reason for generating a warning alert email for low print production is a network outage. If the alert email depends on the external network to be routed to staff, the network outage may go unnoticed until print fulfilment staff notice that the printers are working below capacity.