

Manual

Archivista 2014/IX

Contents

I	Introduction	7	II	ArchivistaBox	24
1	Introduction	8	5	ArchivistaBox: first steps	25
1.1	Welcome to Archivista	8	5.1	Benefits	25
1.2	Notes on the manual	8	5.2	Starting ArchivistaBox	25
1.3	Our address	9	5.3	Updating ArchivistaBox	28
1.4	Previous versions	9	5.4	Notes on the box models	29
1.5	Licensing	14	5.5	Virtualised ArchivistaBoxes	29
2	Installation	16	5.6	ArchivistaBox DOLDER	30
2.1	Physical ArchivistaBox	16	5.7	ArchivistaBox RIGI	30
2.2	Virtualised ArchivistaBox	16	5.8	ArchivistaBox SAENTIS	31
3	Ten steps towards success	17	5.9	ArchivistaBox PILATUS	31
3.1	Step 1: Starting it up	17	5.10	ArchivistaBox TITLIS	31
3.2	Step 2: Accessing the web interface	17	5.11	ArchivistaBox EIGER	31
3.3	Step 3: Upload document	17	5.12	ArchivistaBox MATTERHORN	32
3.4	Step 4: Share FTP/SMB directory	17	5.13	Scan station Albis (optional)	32
3.5	Step 5: Printing documents to the archive	17	5.14	Scan station MYTHEN (optional)	32
3.6	Step 6: Scanning with the WebClient (KeyPad)	17	5.15	Scan station ROTHORN (optional)	33
3.7	Step 7: Setting up a backup facility	17			
3.8	Step 8: Restart ArchivistaBox	18	III	WebClient	34
3.9	Step 9: Image import from camera/USB stick	18	6	WebClient	35
3.10	Step 10: fields and forms in WebAdmin	18	6.1	Login	35
4	After kick-off...	19	6.2	Views	35
4.1	Introduction	19	7	Navigation	36
4.2	ArchivistaBox	19	7.1	Main view	36
4.3	Operating system	20	7.2	Sorting	37
4.4	Virtualisation	20	7.3	Page view	37
4.5	RAM mode and ISO file	20	8	View, Search and Edit	38
4.6	Document management (DMS) terms	20			

8.1	View (Ctrl+F5)	38	12.2	Field type	60
8.2	Search (F5)	38	12.3	Length	60
8.3	Edit mode (Shift+F5)	40	12.4	Position after	60
9	Extended functions	44	13	Mask definition	61
9.1	Uploading a file	44	13.1	Mask	61
9.2	Copy template	44	13.2	Field name	61
9.3	Create PDF file	44	13.3	Field type	61
9.4	Creating and scanning documents	44	13.4	Link with field	63
9.5	Combine documents	45	13.5	Label	64
9.6	Editing and deleting pages	46	13.6	Position	64
9.7	OCR Parameters	46	13.7	Width	64
9.8	Notes	46	13.8	User(s) allowed new entries	64
9.9	Text recognition for page (with notes)	48	13.9	User(s) allowed changes	64
9.10	Download	49	14	Archive administration	65
9.11	Printing pages	49	14.1	Fields in the archive	65
9.12	Barcode processing	49	14.2	Options for page text and search	67
9.13	Versioning	50	14.3	Options for archive	68
9.14	Attributing keywords with F7 and F8	50	14.4	Enabling versioning	70
9.15	Search terms of previous search	50	15	Scan definitions	72
9.16	Replace	51	15.1	Introduction	72
9.17	Scanning documents	51	15.2	Settings	72
IV	WebAdmin	53	15.3	Post editing	74
10	WebAdmin	54	16	Barcodes (Archivista Box)	77
10.1	Login	54	16.1	General	77
10.2	Logout	54	16.2	Barcode technology	77
11	User	55	16.3	Barcode entry	78
11.1	Administration (in general)	55	16.4	Barcode recognition	84
11.2	User administration (external)	57	16.5	Barcode processing	85
12	Field definition	60	17	OCR definitions	86
12.1	Field name	60	17.1	Name of OCR page definition	86
			17.2	Languages of definition	86

17.3 Text quality of pages	86
17.4 Further preparatory settings	86
17.5 Options for table recognition	86
17.6 Time for optical character recognition (OCR)	87
18 SQL definitions	88
19 Form Recognition (ArchivistaBox)	89
19.1 Introduction	89
19.2 Managing forms	89
19.3 Editing objects of currently active definition	90
19.4 Logo recognition	93
19.5 Summary	95
20 Exporting documents	96
20.1 Introduction	96
20.2 Exporting files in WebAdmin	96
20.3 Exporting files in the WebClient	96
21 Mail archiving	97
21.1 Introduction	97
21.2 Individual definitions	97
21.3 Editing a definition	97
21.4 Possible problems during setup	98
22 Administrating jobs	99
22.1 Introduction	99
22.2 Example	99
23 Database creation	100
23.1 Option 'create'	100
23.2 Option 'drop'	100
V WebConfig	101
24 Administration with WebConfig	102

24.1 Introduction	102
24.2 Login	102
24.3 Setup scan button	102
24.4 Backup	103
24.5 Services (CUPS, FTP, remote access)	103
24.6 Unlock documents	104
24.7 Passwords ArchivistaBox	104
24.8 View log files	104
24.9 Text recognition (OCR)	105
24.10 Turn off ArchivistaBox	105

VI Desktop 106

25 Desktop ArchivistaBox 107

25.1 Accessing the system menu	107
25.2 Archivista WebClient (Alt+F10)	107
25.3 Archivista WebAdmin (Alt+F9)	107
25.4 Archivista WebConfig (Alt+F11)	107
25.5 View manuals	107
25.6 Archiving & OCR	108
25.7 Archivista modules	111
25.8 Backup	114
25.9 Encryption	116
25.10 Print server	117
25.11 FTP Server	119
25.12 Mail server	121
25.13 Database	121
25.14 System	126
25.15 Remote Access	131
25.16 Exit (Alt+F4)	132
25.17 Function keys	132

VII	ArchivistaVM	133
26	Server virtualisation with KVM	134
26.1	Introduction	134
26.2	System requirements	134
26.3	Login to ArchivistaVM	135
26.4	System overview	136
26.5	VM Manager	136
26.6	Configuration	141
26.7	Administration	145
26.8	Cluster	146
26.9	Working with the console	148
26.10	Repairing hard disk networks	150
26.11	ArchivistaVM-Server	151
VIII	ArchivistaERP	152
27	ArchivistaERP	153
27.1	Introduction	153
27.2	Activate ERP system	153
27.3	First Steps	155
27.4	Creating an invoice	156
27.5	Sales	157
27.6	Purchases	157
27.7	Items and Inventory	158
27.8	Manufacturing	158
27.9	Dimensions	158
27.10	Banking and General Ledger	158
27.11	Setup	158
IX	Publishing	159
28	Publishing Edition III	160
28.1	Long-term considerations on documents	160

28.2	The Schneider adventure	160
28.3	The XP adventure	160
28.4	Why source formats are not suitable	161
28.5	Preliminary note	162
28.6	Publishing self-supporting archives	162
28.7	A few tips on working with documents	164

X Internals 168

29 External SQL calls (API - WebClient) 169

29.1	External login	169
29.2	Extended access	169
29.3	Examples of working with API Interface	178

30 Database structures 183

30.1	The most important Archivista tables	183
30.2	Archiving concept	183
30.3	Clean login to Archivista databases	187
30.4	File name of an external saved page	187
30.5	Field names	188

XI FAQs 190

31 ArchivistaBox 191

31.1	Is there a demo version of Archivista?	191
31.2	Is ArchivistaBox free?	191
31.3	ArchivistaBox does not start up properly	191
31.4	Shutdown of ArchivistaBox	191
31.5	Installation/Update from CD	191
31.6	Installing ArchivistaBox on a Virtual Machine	192
31.7	Saving your data	192
31.8	Data exchange and backup by means of USB stick	192

32 Integration questions 195

32.1 Is there a mail archiving function?	195	
32.2 Can I archive Office files (incl. check out)?	195	
32.3 Which possibilities are there with regard to scan- ning to ArchivistaBox?	195	
32.4 Can I scan books?	195	
32.5 Classifying documents (ERP interface)	196	
32.6 Is there an API for the ArchivistaBox?	197	
32.7 Import of images	197	
32.8 Transferring the RichClient to the ArchivistaBox . .	199	
		32.9 Importing data from other DMS software 199
33 Solutions for common problems	200	
33.1 Scanning is slow	200	
33.2 Scanner is not working	200	
33.3 Characters on keyboard do not work	200	
33.4 OCR is not (or no longer) functioning	200	
33.5 ArchivistaBox is no longer working	201	
33.6 A user has forgotten their password	201	
33.7 'Document ... locked' message is displayed	201	

Part I

Introduction

1 Introduction

1.1 Welcome to Archivista

Time flies: Quite a few years have passed since the last major release. In fact, for more than two years there have not been any major release versions. Thanks to automated processes and an avoidance of external dependencies where possible (ArchivistaBox is being developed and manufactured in Switzerland) we can develop our products continually and steadily and release them at shorter intervals. In principle, there are no major releases anymore, there are a greater number of smaller 'editions' instead.

And yet, I would like to single out in this preface one particular feature of the last years: the **RAM mode**. **All our products run completely in primary storage**, only the data is saved on non-volatile memory disks. Why is this so vital for a product such as ArchivistaBox?

Firstly, a software running in primary storage is no longer available when the computer has been shut down. In other words: the software must be set up again completely when the computer is started. Those who work with operating systems and specialised applications daily (or often) know that setting up software correctly is extraordinarily time consuming. This is not so in the case of ArchivistaBox. By copying it from one single file (ISO) to the memory in completely automated fashion, the process is highly efficient. **After far less than one minute an ArchivistaBox is set up and ready to work with.**

To cut a long story short, an ArchivistaBox does not need to be installed and set up anymore. It is simply launched and put immediately at the disposal of the user. On subsequent updates it is sufficient to run a new file (ISO) and the new release is available. The entire process is automated. This means that no IP data

or other settings need to be entered manually. In fact, customer specific data is entered once and committed henceforth together with the release. If a customer works with several ArchivistaBoxes (e.g. ArchivistaDMS with several scan boxes), **the same release (ISO) can be used for all ArchivistaBoxes. The process governing the start-up recognises its role automatically on the basis of the initially entered specifications and installs itself accordingly.**

For all of these reasons we speak with regard to **ArchivistaBox of zero maintenance; there are no installations nor updates, there are simply newer or older versions and these can be get up and running.** This is also why chapter 'Installation' seems alarmingly short at first sight. It may happen that even experienced IT specialists are at a loss when they engage in documenting the installation and update process - maybe strictly following company policy - and 'fail' quite simply because there is nothing to document. In this sense, we share our happiness with you that ArchivistaBox has become so straightforward and uncomplicated. All the same, we will not rest and we will keep aiming at increasing ArchivistaBox' performance and ease of use. We look forward to continued, successful cooperation with you also in the future.

Urs Pfister, Managing Director, Archivista GmbH

1.2 Notes on the manual

This manual is available on paper and as an online **help file**. The latter can be started directly from within the program with the **function key 'F1'**.

➡ You can do a full text search on the online help file. Just enter your search string in the field at the top or the bottom of the dialog box and click on 'Search'.

Note on upgrade from earlier versions: From 1 January 2014 the RichClient application as well as ArchivistaBox 32Bit are no longer available. Instructions about how to upgrade from these versions can be found in the documentation of earlier versions.

➡ Important notice or tip: whenever you see this symbol, you get relevant additional help.

1.3 Our address

You can reach us under the following address:

Archivista GmbH, Stegstrasse 14, CH-8132 Egg,

Tel: +41 (0)44 350 05 60, Fax: +41 (0)44 350 05 61,

E-Mail: webmaster@archivista.ch.

1.4 Previous versions

1.4.1 Overview

From the mid-nineties when the Archivista product suite came into being a number of versions have been released. In chronological order:

1998: RichClient/RichAdmin: The first version of Archivista runs under all Windows versions.

2002: WebClient: Endusers no longer need to retrieve and view documents via an installed program but can do it by means of a web browser. Advantage: greater flexibility, less implementation effort. Document editing and administration tasks are done with the RichClient.

2005: Webadmin: Now endusers can also add documents and edit records via the web browser. Also administrators can make settings by means of a browser based interface.

2006: ArchivistaBox: WebClient and WebAdmin run on mini archiving servers that are pre-configured. In addition, all functions relating to the administration of the archives (backup, encryption, remote access, etc.) are integrated in the ArchivistaBox.

2007: ArchivistaBox with AJAX (Web2): Ordinary web applications may be slower than RichClient applications because of the form concept. Thanks to new web technologies (AJAX in particular) it is possible to display content dynamically without having to reload forms. This is the reason why this technology is introduced with ArchivistaBox 2007/III. With version ArchivistaBox 2007/VIII also the RichClient is 100 percent OpenSource.

2008: ArchivistaBox as a BusinessBox: In addition to the web-based WebConfig application (web-based system settings), there is now an optional ArchivistaERP module. This now makes it possible to use the ArchivistaBox as an ERP Box, too. Furthermore, the Linux port from Cuneiform provides good Open Source recognition which can also create searchable PDF files.

2009: With the integration of Office formats over 200 file types can now be processed directly. Documents archived in this way can then be checked out again (in their original versions). In addition, a more powerful mail archiving function is provided. WebClient and the ArchivistaBoxes, Albis and Matterhorn, are also new.

2010: Many new features have been added with versioning, API and WinUpload. Documents can now be automatically provided with a version stamp. With the optional API module, the WebClient can be controlled remotely, and files can be easily uploaded to the WebClient from Windows using the API module.

2011: ArchivistaBox 64-bit. After more than 5 years, we have migrated ArchivistaBox to 64-bit. ArchivistaVM is now available as a virtualisation module. A new setup program makes it possible to set up ArchivistaBox in 1 to 2 minutes.

2012: ArchivistaRAM and cluster By and by all ArchivistaBox versions are converted to RAM mode. This means that installations have become unnecessary. The system becomes operational directly from the ISO file. In the case of ArchivistaVM there are cluster versions, i.e. hard disk content can now be copied to two computers simultaneously (including support for 10 GBit).

2013: Scan boxes and WebClient In the year 2013 extremely fast document scanners were implemented (far above 100 pages per minute), now text recognition can be completed in parallel. With Albis III there is a new scanbox – weighing ca 100 grams – that processes up to 200 pages per minute. The WebClient comes with column support for the fields; thus the RichClient can definitively be replaced.

2014: ArchivistaBox Dolder With ArchivistaBox Dolder a DMS can be realised at even lower cost.

1.4.2 ArchivistaBox 2014/III

At the same price ArchivistaBox Dolder is supplied with hardware. Furthermore, the most current scanner drivers are available on the ArchivistaBox systems. ArchivistaVM allows the user to conveniently extract data via the web interface.

1.4.3 ArchivistaBox 2014/I

Ansi export from the WebClient, restoring of guests directly in ArchivistaVM (web interface). Simpler offline update (including instructions for Windows, to create a bootable USB stick). For the sake of simplification ArchivistaDesktop is no longer available

since ArchivistaDesktop used to be identical to ArchivistaDMS (with the exception of a few additional office applications).

1.4.4 ArchivistaBox 2013/X

Fully renovated WebClient, which enables the user to display up to three columns in the main view. On top of that, the WebClient adapts to screens of different sizes in an optimum way (from portable phone to 40 inch screen).

1.4.5 ArchivistaBox 2013/V

With the new scan station Albis III there is a scan box available that weighs only 100 grams and scans up to 200 pages per minute.

1.4.6 ArchivistaBox 2013/II

The new Archivista box systems contain the drivers for the very fastest document scanners (processing far above 100 pages). In order that no backlog may be accumulated, OCR can be processed in parallel.

1.4.7 ArchivistaBox 2012/XI

The ISO files were drastically reduced. ArchivistaBox can be started from CD, memory stick, network (PXE), internal hard disk and external backup disk. ArchivistaDMS administrates master/slave instances automatically. Equally and massively simplified was the creation of archive folders.

1.4.8 ArchivistaBox 2012/IX

Newly implemented in the WebClient is a Search/Replace feature. Further, the current search values can be saved and values of any document can be used as template for efficient categorisation by keyword with a click.

1.4.9 ArchivistaBox 2012/VII

For the first time ArchivistaBox can be started completely from the main memory (RAM mode). Equally new is the on-the-fly publication of archives; here too the RAM mode is used. ArchivistaVM comes with user administration.

1.4.10 ArchivistaBox 2012/II

ArchivistaBox is newly refurbished (new Linux basic system).

1.4.11 ArchivistaBox (first generation)

1.4.11.1 ArchivistaBox 2012/I

With the help of a new build concept all IP data per customer is captured. Subsequently, the software can be updated anytime; in automated and customer-specific fashion.

1.4.11.2 ArchivistaBox 2011/VI

ArchivistaERP is released with a new version supporting ESR (Swiss payment orders). WebConfig is set up afresh in order that settings may also be modified on a slave computer.

1.4.11.3 ArchivistaBox 2011/V

ArchivistaVM supports DRBD. Hard disk content can be mirrored to two different computers at the same time. With the help of fast 10Bit cards and disk arrays of up to 24 hard disks extremely fast hard disk clusters can be realised.

1.4.11.4 ArchivistaBox 2011/IV

ArchivistaBox 2011/IV is the first 64-bit ArchivistaBox: It is available in three versions: ArchivistaVM (ca. 330 MB) to deliver Server Virtualisation (KVM) to the desktop, ArchivistaDMS (ca.

700 MB), which includes both ArchivistaBox and ArchivistaVM, ArchivistaDesktop (ca. 920 MB), which in addition to virtualisation and document management (DMS) includes other desktop applications such as Scribus and Gimp. The new ArchivistaBox 64-Bit can be set up in 1 to 2 minutes, and the setup now supports hard drive arrays (software RAID).

1.4.11.5 ArchivistaBox 2010/IV

For the first time, a versioning function is available for ArchivistaBox 2010/IV. Versioning (available in all ArchivistaBox systems) can (automatically) monitor changes to Office, PDF and/or other image files uploaded to the ArchivistaBox.

1.4.11.6 ArchivistaBox 2010/III

Note field function for ArchivistaBox archives has been added. Scripts can now be managed centrally in WebAdmin for several ArchivistaBoxes. A code (rather than a fixed number) can now be assigned to scan definitions. Keyword assignment can be automated with PDF Creator (Windows). Furthermore, the winupload.pl tool can be used to upload files easily via the API module of the WebClient.

1.4.11.7 ArchivistaBox 2009/VII

A new WebClient offers the user a more intuitive interface. Table views can be enlarged and reduced at will, sticky notes can be added. The newly integrated free text engine Sphinx offers support for archives containing well over 100 million pages. With scan box Albis there is a scan solution that can be configured conveniently by means of a USB stick, and ArchivistaBox Matterhorn allows management of up to several TB (sufficient space for 50 to 100 million pages with a standard ArchivistaBox).

1.4.11.8 ArchivistaBox 2009/I

The ArchivistaBox 2009/I can be used to send all Office documents (OpenOffice and Microsoft Office) to the archive (a function previously reserved for the old RichClient). Documents archived in this way can then be checked out again (in their original versions!) to undergo further processing. All formats (over 200 file types) can be processed without the need for external plug-ins. A more powerful mail archiving function is also provided which can be used to archive entire IMAP folders automatically. The system menu can also be used to update the ArchivistaBox via the Internet. Split archive tables now make it possible to create archives of many terabytes in size.

1.4.11.9 ArchivistaBox 2008/XI

An optional LCD panel can now be ordered which can be used to trigger scanning via a small display. All settings can also be called from this display. In addition, the ArchivistaBox 2008/XI can now be run on netbooks (tested with the EeePC series). A new mobile scanner and the non-virtualised variant of the Dolder ArchivistaBox complement this perfectly.

1.4.11.10 ArchivistaBox 2008/IX

New Eiger ArchivistaBoxes are being introduced with the ArchivistaBox 2008/IX. On average, these are two to four times more powerful than the old Eiger Boxes. Good Open Source text recognition (Linux port from Cuneiform) is now provided as well. In addition to text recognition, this can be used to create searchable PDF files in around 20 languages.

1.4.11.11 ArchivistaBox 2008/VI

French version: The WebClient, WebAdmin and the system interface are available in French. A new administration application for

the system settings is provided with Archivista WebConfig. A new file upload is also supplied. The new API interface for the WebClient will probably also interest programmers. The web-based ArchivistaERP application is also new.

1.4.11.12 ArchivistaBox 2007/VIII

Beside various minor improvements there is now a comfortable import feature for images coming from digital cameras or USB sticks, respectively. Another new feature is form recognition which - together with logo recognition - enables the user to process documents automatically. Furthermore, the entire source code under `/home/cvs/archivista` is now subject to the GPL license version 2. Lastly, ArchivistaBox can be attached directly to an LDAP server for central user management.

1.4.11.13 ArchivistaBox 2007/III

Totally new WebClient (Ajax based), improved scan engine (parallel processing of scan jobs), faster OCR on ArchivistaBox. Merging of documents (RichClient). Archived folders can be automatically written to CD/DVDs. All sessions can be protocolled in detail (WebClient) and viewed by means of the application AccessLog.

1.4.11.14 ArchivistaBox 2006/XI

Black/white optimisation for badly printed sheets of paper, splitting of documents (even without barcodes) during scanning process, automated writing of DVD/CDRW with complete archives, remote access for easy maintenance, multi core processor support with up to 4 GB RAM, automated creation of archives on USB sticks.

1.4.11.15 ArchivistaBox 2006/III

First released ArchivistaBox edition: print and FTP server. Back-ups on USB stick, tape, Windows/Linux file systems and USB hard disks. Master/Slave configurations. PDF/OCR server and barcode recognition module.

With ArchivistaBox 2006/III a new versioning concept begins. The year takes the first position followed by the month (in Roman numerals). The software of the ArchivistaBoxes are subject to the GPL, i.e. the source code is open.

1.4.12 RichClient (support expired)

1.4.12.1 Archivista 5.2

From **version 5.2** onwards there is not only the RichClient version of Archivista but also the embedded OpenSource solution ArchivistaBox RIGI, PILATUS, TITLIS and EIGER. Smaller innovations include a photo mode for photo archives and an editor for templates in the barcode print program (1 Oct 2005.)

1.4.12.2 Archivista 5.1

Version 5.1 offers many new features apart from a web module and a PublishingEdition for self-supporting archives. (15 April 2004).

1.4.12.3 Archivista 5.0

Version 5.0 embeds MySQL as database server and is able to handle much larger archives than hitherto. (30 March 2003).

1.4.12.4 Archivista 4.1

Version 4.16c works with FineReader5 engine (incl. PDF export) and offers post-editing functions for scanned documents (8 May 2002).

Version 4.15 introduces a fourth archiving format (JPeg). The archiving format may be specified on document level (15 January 2001).

Version 4.14 contains an enhanced FineReader4 engine and the new Search33 engine to index even larger archives. In addition to that there are many further enhancements (16 June 2000).

Version 4.12: FineReader 4.0 technology is introduced for OCR and a comprehensive barcode processing solution is incorporated too (19 July 1999).

Version 4.11 introduces two additional file formats. Along with that an English version is issued for the first time (30 Dec 1998).

1.4.12.5 Archivista 4.0

Version 4.0i was completely refurbished. In contrast to earlier versions Archivista 4.0i is no longer available as shareware version. (26 May 1998).

1.4.12.6 Archivista 3.x

Version 3.x with new modules 'Archive', 'Adresses', 'Notes' and 'Literature'. Version 3.x runs under Windows 9x/NT (1 June 1997).

1.4.12.7 Archivista 2.x

Version 2.x runs under Windows 3.x and Windows9x/NT (30 August 1996).

1.4.12.8 Archivista 1.x

Version 1.x: Internal version (20 January 1996).

1.4.13 Nobody is perfect

Archivista was developed over a period of more than fifteen years. Each version is amply tested before it is released. Nevertheless: nobody is perfect! Archivista GmbH is grateful for any indications regarding bugs or inconsistencies in the program functions.

1.5 Licensing

1.5.1 Sources, offers free of cost and ArchivistaBox

The applications WebClient, WebAdmin and WebConfig, ArchivistaERP and ArchivistaVM are Open Source (GPLv2 licence). The sources can be found here:

www.archivista.ch/de/media/archivista-gpl.tgz

In addition, there are the following offers free of cost: ArchivistaCloud (cf. 4.2.3) and ArchivistaBox Mini (cf. 4.2.4). These you may operate in the framework of the private use rule. You may use these products too, if you have previously purchased an ArchivistaBox; in this way you get test environments and will be enabled to work with ArchivistaBox (even) more effectively.

Unlike the ArchivistaBox sources, which with a few exceptions (e.g. commercial text recognition) are subject to the GPL license, this is not and was never the case for the ISO files, the manual, the Archivista website content or the logos. That is why you may not copy, modify or distribute the ISO files, the manual, the Archivista website content nor our logos. You can purchase ArchivistaBox exclusively from one of our certified distribution partners or from Archivista GmbH. This always includes a maintenance contract (for the first year it is included in the basic price), cf. also 1.5.3.

In addition, please note that Archivista is a registered word/figurative trademark. It is therefore not permitted to use the word/figurative trademark nor the entire ArchivistaBox in any

other form than that used in the unmodified ISO file(s), which is true to the motto, that if ArchivistaBox is on the box, it is also inside the box.

🔑 **Warranty:** We guarantee you that all distribution partners listed under <http://www.archivista.ch/en/pages/ordering.php> are official distribution partners for ArchivistaBox. They are authorised to implement the solution for you. Each distributor has at least one ArchivistaBox and has completed training with us. No other persons/parties are entitled to distribute ArchivistaBox commercially.

1.5.2 Copyright note

Archivista GmbH, Stegstrasse 14, CH-8132 Egg possesses all rights regarding its software 2014/IX .

Trademarks: Postscript & Acrobat (Adobe), Microsoft Word, Excel, Access and Windows (Microsoft), Apple Macintosh (Apple), Zweckform (Zweckform Buero Produkte GmgH), Ghostscript (Artifex) and Ghostgum (Ghostgum), AXIS (Axis Communications) and Xerox (Xerox Corporation). ABBYY(TM) FineReader(TM) OCR © 1993-2006 by ABBYY Software House. All rights reserved. ABBYY, FineReader are trademarks of ABBYY Software House. Barcode from Softek Software www.softeksoftware.co.uk. VMware is a registered trademark of VMware, Inc., USA. www.vmware.com. MySQL, VirtualBox and OpenOffice are registered trademarks of Oracle, Inc., USA.

All Archivista applications 2014/IX are protected by Swiss copyright law and international treaties. VirtualBox and OpenOffice are registered trademarks of Oracle, Inc., USA.

1.5.3 Support for the ArchivistaBox

When you purchase an ArchivistaBox, a maintenance contract for the first year is included in the price. This includes guaranteed

response time to faults, a free replacement box in the event of a hardware defect as well as all updates for the duration of the maintenance contract. You can obtain support from the supplier of the ArchivistaBox, i.e. the distributor from which you have purchased the ArchivistaBox. Should a supplier no longer be among the ArchivistaBox distribution partners, you will receive support either from another distributor or directly from Archivista GmbH. In this case, you agree to continue the service contract seamlessly within its normal scope with the new party.

The maintenance contract always runs (provided no other arrangement is agreed in writing at the time of the purchase of the ArchivistaBox) for one year and is automatically extended for a further period of one year, unless the purchaser or the supplier ends the maintenance agreement by giving notice of at least three months before the end of the respective support year. A change in terms and conditions (e.g. price adjustments) by the supplier can be communicated using the same notice period.

1.5.4 Guarantee

Archivista GmbH guarantees that the application Archivista 2014/IX will run in basic accordance with this documentation for 90 days after selling date. The total liability of Archivista GmbH consists of a refund of the originally paid price or of subsequent improvement of the application (the choice of which is Archivista GmbH's). Misusing the program or manipulating it in a wrong manner voids this guarantee. There is no further guarantee. In particular, any liability with regard to consequential damages is excluded (as far as is lawful). Swiss law applies and the place of jurisdiction is Zurich.

2 Installation

2.1 Physical ArchivistaBox

You have bought an ArchivistaBox. It sits in front of you and you would like to start it up. Since all IP data has been gathered before the delivery of the box, you can simply turn it on. Before actually doing it, you should connect ArchivistaBox to your network by means of a network cable. Further information on your ArchivistaBox can be found at [II](#)

2.2 Virtualised ArchivistaBox

You have bought an ArchivistaBox too but you would like to operate it virtually. In this case you have received an installation CD. Generate a model of ArchivistaBox according to the basic data of the purchased version. You can subsequently start ArchivistaBox, the system will set itself up autonomously.

➡ Along with the information of where you can buy ArchivistaBox you will receive (a) valid MAC address(es) for the ArchivistaBox(es): You must enter it/them next to the network interface card when defining the virtualised instance. Should you want to utilise a specific MAC address, you can mention it when you place the order. In this case the Archivista ISO file will be created using the Mac address provided by you.

3 Ten steps towards success

3.1 Step 1: Starting it up

Connect ArchivistaBox with network, attach keypad and scanner (if the box is used for scanning purposes). Turn on ArchivistaBox, cf. 2.

3.2 Step 2: Accessing the web interface

After one minute at the latest ArchivistaBox must be visible in the network. Enter the IP address on a different computer and you should get to the login dialog of ArchivistaBox. It may be that the ArchivistaBox certificate must be linked to the browser. More information on accessing the WebClient can be found at 5.2.2.

3.3 Step 3: Upload document

After login (initial password for 'Admin' is 'archivista'), you should test document upload manually. This procedure is described at 9.1.

3.4 Step 4: Share FTP/SMB directory

Now you can give permission to the FTP service in WebConfig, cf. 24.5 and 25.11. Subsequently, you can upload documents via FTP (user name is always ftp) and via SMB. Office documents for the database 'archivista' are filed in subfolder 'office/archivista'. In the same way you could integrate documents by means of a network scanner in folder 'pdf/archivista' (text recognition is triggered automatically). The format of the folders and subfolders is described in 25.11.1.1.

3.5 Step 5: Printing documents to the archive

To print documents to the archive you need to activate the print server of ArchivistaBox and correspondingly create a printer on client side (e.g. Windows). This is described in 24.5 and 25.10.

3.6 Step 6: Scanning with the WebClient (KeyPad)

If you have connected a scanner to the ArchivistaBox, you can start scanning with the help of the WebClient. More information can be found in 9.4.

➡ If a keypad and a scanner are attached you can trigger scanning by means of the keypad. On first start-up there are two definitions: B/W and colour. With 1+Enter you scan in B/W, with 2+Enter in colour. Further information can be found in 9.17.2.

Please note that this option is only available if you have purchased a keypad and a scanner driver for your ArchivistaBox.

3.7 Step 7: Setting up a backup facility

Define how backups should be run in WebConfig. Backups should be carried out periodically. More information can be found in 24.4. Please take note that there is a feature called **Restore-on-the-Fly**, too. More information on this feature can be found in 25.8.3.1.

3.8 Step 8: Restart ArchivistaBox

The changed system settings such as FTP/SMP, print server and/or data backup are saved only after a restart of ArchivistaBox. This process is carried into execution in WebConfig, cf. 24.10.

➡ Alternatively, you can prepare a USB stick by creating a folder named 'config' and attach it to the ArchivistaBox in order to save the system settings. Cf. .

3.9 Step 9: Image import from camera/USB stick

If you would like to import images from a camera or from a USB stick to ArchivistaBox, then take a look at 32.7. There is an application of its own for these cases.

3.10 Step 10: fields and forms in WebAdmin

Now you can define an archive structure that is convenient to you. WebAdmin is the tool for this. Cf. 13.

4 After kick-off...

4.1 Introduction

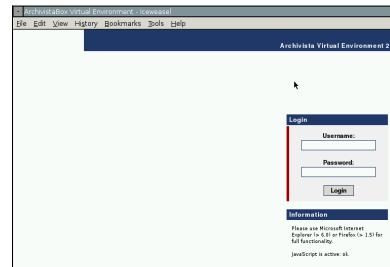
The following introduction provides you with some useful tips for getting started with an ArchivistaBox. If you want to start using the ArchivistaBox immediately, you will find the necessary information under 2.

4.2 ArchivistaBox

ArchivistaBox represents a virtualisation and document management solution (DMS). In this sense, ArchivistaBox is a software solution. Unlike many other solutions, however, the ArchivistaBox is delivered as a complete solution (hardware, operating system and application software). The customer thus receives a complete solution from one source.

4.2.1 ArchivistaVM

Whenever ArchivistaVM is mentioned, it refers to the server virtualisation available on every ArchivistaBox 64Bit. ArchivistaVM means primarily the ArchivistaVM installation CD. Unlike ArchivistaDMS, the ArchivistaVM CD only contains the virtualisation software; the document management application module is not included on it.



The above illustration shows the ArchivistaVM server after starting ArchivistaBox.

4.2.2 ArchivistaDMS

ArchivistaDMS represents the document management system (DMS) of ArchivistaBox. This installation CD is usually shipped together with ArchivistaBox systems.



The above illustration shows the ArchivistaDMS server after starting the ArchivistaBox. ArchivistaDMS always contains the module ArchivistaERP.

4.2.3 ArchivistaBox Cloud

ArchivistaBox Cloud is an opportunity to run ArchivistaDMS and ArchivistaERP on the servers of Archivista GmbH free of cost and for private usage. With purchasing ArchivistaBox you are invited to create a cloud instance. You can, for example, test new features or play around with fields and views without having to install anything.

4.2.4 ArchivistaBox Mini

ArchivistaBox Mini is a tiny virtualisation platform (currently 60 MByte). With the purchase of an ArchivistaBox you are invited to obtain an ArchivistaBox Mini, for example, to test published archives.

4.3 Operating system

To work with a computer, you need software to switch the computer on and off and to install software or programs. To make this possible, each computer needs an operating system. With ArchivistaBox the operating system comes bundled with the software.

4.4 Virtualisation

Instead of installing the operating system directly on a computer (server), a management application is installed first. This software is then used to install guest operating systems. These guest systems are made to "believe" that they have been installed on a physical computer. The virtualisation software "pretends" to be the whole computer, making available everything one would expect from a normal computer, including processors, memory, hard drives, and screen.

4.5 RAM mode and ISO file

The RAM mode (the entire system runs in the RAM, only the data is located on the hard disk) is described in detail in chapter 'Introduction'. We recommend reading 1.1.

The ArchivistaBox systems are pre-configured, i.e. no installation is needed. In spite of this, all customers have access to the

corresponding ISO file (installation file). You can use this file to boot ArchivistaBox from a CD ROM drive or a USB stick. How to create the USB stick is described at length on our website www.archivista.ch.

4.6 Document management (DMS) terms

4.6.1 The digital archive

From the time when information was first available in writing or in the form of pictures the problem arose that knowledge thus collected would not be available when it was needed. This is where archiving or a document management system (DMS) sets in.

In contrast to a conventional filing system where every single page must be filed into folders or cabinets to be extracted later in a time-consuming search, Archivista helps you to do this work in a fraction of the time and in fully digitized manner.

This **time gain** is the result of well thought out processes of entering the material and searching for it effectively. Integrated **full text retrieval**, automated **barcode processing**, individual **keyword indexing**: All these functions are available to you in order that any document you are looking for appears on your screen within seconds.

Document management or digital archiving with Archivista means further that you can store both **paper documents** and documents that exist already in the computer as **electronic files**. Archivista is geared towards making information **available for years** – even **decades** – exactly in the same form as it existed in the original.

To effect this, all documents are stored in a picture format (before being compressed). Such **picture formats (bitmaps)** are very simple and therefore especially durable. Archivista works with **open and freely available standard bitmap formats**.

It goes without saying that we could file all our documents in a tree directory on our hard disk. With a small archive this would work well. But the more documents we have in our archive the more difficult the administration of our data would be. This is the reason why Archivista 2014/IX works with a high-performance database.

4.6.2 Database, server and client

At this point we would like to clarify a few terms whose understanding makes working with Archivista easier.

➡ If these terms are clear to you, please go directly to chapter 4.6.4.

Database: Organizational unit whose purpose it is to collect and retrieve information on the basis of particular criteria. A database could also be likened to a strong room where data are securely managed and stored.

Server and clients: In order to keep our data safe we choose a safe location (server). This location may only be accessed by customers (clients) who have identified themselves as being rightful customers. To put it in a more abstract manner: the server offers clients services; in our case the service is a database.

Information exchange takes place in small quantities. If the customer (client) asks for the whole content of the strong room, he gets the number of all available objects and the content of the first ten objects, but not the 990 other objects. If he wants to see more objects he must ask for objects 11-20. This helps to keep the strong room (server) uncluttered and make the exchange of data more efficient.

4.6.3 Tables, records and fields

The basic principle is: In the case of ArchivistaBox each database corresponds to one archive. Several databases (archives) can be

managed. A database contains tables, these contain records which themselves contain fields (reappearing in each record) and in the fields a specific piece of information is stored (e.g. a number).

Tables: The function of tables consists of pulling together the same type of information and making it visible (e.g. 'documents' in table 'Archive'). In Archivista there are other tables, e.g. table 'Archive images'.

Record: Each record (corresponds to 'document' in table 'Archive') contains a specified (identical) number of fields.

Fields: The field is the smallest organization unit and houses one piece of information (e.g. the date of the document).

4.6.4 Archivista and working method

Now that you are ready to take the plunge into the world of electronic archiving let us make a few remarks regarding Archivista and working method. Probably you have worked with folders hitherto, i.e. each document was filed in a folder. In the course of time you developed a folder structure. You might have had a separate folder for orders, invoices and reminders, for example. Alternatively, it may have been your clients which were the organizational unit of your filing system.

This **thinking in terms of folders is very inefficient from the point of view of digital archiving**. It simply does not make any sense to haul folders if all you are interested in is one document. Looking for a document may be a very time consuming task because neighboring documents not only guide your search but hamper it too. An index helps structuring your material but if you are forced to extend it, an often complicated reorganization follows suit.

That is the reason why we at Archivista GmbH have deliberately decided against such time consuming folder structures. You will not find any imitated safes, cabinets, folders and registers

in Archivista because their disadvantages in the realm of digital archiving exceed their advantages by far. Give your documents the information according to which you will search for them later. An invoice is named invoice and by adding a client name or client number and a date your document is uniquely defined. That is all. When you look for your documents later on you do not have to dig in your archive for the appropriate folder but you simply tell Archivista that you want to see all invoices addressed to a specific client and that you want to see them sorted according to date. Within seconds you get a list of documents which correspond to these criteria. This is much faster than working your way through filing cabinets, folders and masses of documents in the same sub-folder.

Do not only throw your old, dusty archive overboard but the working method that goes with it, too. Move on to the new advantages which digital archiving with Archivista offers you. Enjoy!

4.6.5 A few thoughts on document management systems

What number of documents do you have? What sort of documents are they? The higher the number the more sense does a powerful scanner (including automatic document feeder) make. Apart from the velocity of the scanner the quality of the document feeder is highly important. Does it handle your specific (thin or thick) sheets without any problems? We gladly come up with suggestions for scanners which fulfill your particular needs. Please send an e-mail to webmaster@archivista.ch.

If your filing system relies on **simple categories** you must decide whether manual **keyword indexing** makes sense or whether barcodes are a good idea. Manual keyword indexing is suitable for up to 100 pages per day. If you have more paper to scan and archive

you should think about barcode processing because barcode processing enables you to allocate keywords to your documents automatically. Thus all your documents are uniquely identified and they can be retrieved with a 100% certainty.

Full text archive: If you scan many texts, the features text recognition and full text retrieval of Archivista are a boon. Do pay attention to choosing the right language set. Umlauts are no problem then and thanks to the extensive dictionaries you may achieve excellent results.

Integration of documents already existing in digital form: Documents which already exist in electronic form may be incorporated into your archive by using an Archivista printer. We recommend that you make yourself familiar with the basic functionality of Archivista before you try a hand at the Archivista printer. Please read chapter 24.5.

Die **Datensicherung** ist und bleibt ein ganz zentraler Punkt bei einem jeden DMS. Die ArchivistaBox bietet vielfältige Möglichkeiten, die Daten zu sichern. Im einzelnen sind dies: a) Sicherung auf USB-Platte oder übers Netzwerk, b) Erstellen von Archivdatenträgern (CD/DVD), c) Publizieren von Archiven sowie d) Restore-on-the-fly.

Backups are and remain vital in archiving and with any DMS. ArchivistaBox offers manifold ways to backup the data. In particular they are: a) storage to external hard disk (USB) or to a directory in the network, b) copying of data to write-once storage media (CD/DVD), c) publication of archives and d) restore-on-the-fly.

4.6.6 Archive, pages and documents

After installing Archivista the first **archive with the name 'archivista'** is created automatically. You can begin feeding

it with your papers straight away. There are two terms to know which is helpful: **documents and pages**.

Each sheet of paper corresponds to a **page**. Documents already existing in your computer are tied into the archive as if they had been printed, i.e. they are sent to a virtual printer and added to the archive.

A **document contains several pages** which belong together. A document may consist of only a single page but each page is allocated to a document.

➡ Each Archivista archive is made up of an undetermined number of documents. Each document consists of one or several pages.

4.6.7 The Archivista document

Each record in table 'Archive' is treated as a '**document**'. The **document number** (field 'Document') is allocated automatically. Each 'document' contains 640 pages at the most and will be assigned to a **folder**. The **folder number** (field 'folder') is generated automatically. At the beginning you work with folder number '1'. Later, as time goes by, the folder number is automatically increased by 1 each time a folder's capacity is reached.

The fields 'Document', 'Pages' and 'Folder' cannot be edited. The same applies to field 'Archived', which indicates whether a document has already been archived. If it has, no more pages can be added.

Part II

ArchivistaBox

5 ArchivistaBox: first steps

ArchivistaBox is a preconfigured server for daily business operations. It is based around document management (DMS) and virtualisation with KVM. In the latter case we are referring to ArchivistaVM, see also 26 below. ArchivistaBox also contains a simple small ERP solution. All user interfaces are web-based, the complete solution is delivered ready-to-use on a single box and the source codes are subject to the GPL, which is to say that the software is available as open source.

The main models are as follows: DOLDER, RIGI, SAENTIS, PILATUS, TITLIS, EIGER and MATTERHORN. In addition, there are three (optional) models which are used as intelligent network scanning stations: ALBIS, MYTHEN and ROTHORN.

➡: If you are only interested in ArchivistaVM (server virtualisation), you can rely on the following powerful models: Budget, Summit and Universal. The information can be found in our web shop at shop.archivista.ch. These models contain only ArchivistaVM; the entire software for the ArchivistaBox is not included in this version. Below we present the individual models of ArchivistaBox (DMS section).

5.1 Benefits

- No need to install it on a server already occupied by other programs or onto client workstations.
- Very intuitive user interface. Hardly any training necessary.
- Documents are available company-wide from a central server; long-term access is ensured by raster image archiving.

- ArchivistaBox is independent of operating systems already in use: it works with Windows, Unix and MacOS/X.
- Installation is simple: Plug in USB flash drive, start installation, set network data (IP address), after 1 to 2 minutes ArchivistaBox is ready to use.

5.2 Starting ArchivistaBox

ArchivistaBox systems are preconfigured, i.e. you do not need to configure them: Connect them to the 'outside world' (network) and you are all set. After attaching the network cable, press the button 'Power' at the front of the ArchivistaBox computer. The box starts up and after approximately 20 to 30 seconds you will see the login screen of the Archivista web client in front of you.

➡ ArchivistaBox Dolder and the scanning station Albis do not have a power button. Simply plug these models into the mains power supply: The boxes will start up automatically.

➡ **Important:** ArchivistaBox Dolder does not allow for attaching a monitor. See 5.6.1. After approximately 1 minute you can conveniently go through the following steps using a different computer.

5.2.1 Structure of ArchivistaBox

There are five operating modes.

- Home: switching between virtualisation and DMS. ArchivistaVM is described in Chapter 26
- WebClient: working as end user (example: adding documents)

- WebAdmin: making modifications to an archive database (example: editing field definitions)
- WebConfig: making modifications to ArchivistaBox itself (example: network address)
- System menu (desktop): access by means of right mouse click and 'Quit' and yet another right mouse click (only available on box)

If you want to work with ArchivistaERP, please take heed of the information in 27.2.

In the following we show you how to access the different operation modes and what they are for.

➡ Section 3 gives you first assistance regarding how to set up ArchivistaBox in your environment.

5.2.2 Login to WebClient

After ArchivistaBox has been switched on, the WebClient login dialog appears. With the WebClient you can create and modify documents.

[WebClient](#) - [WebAdmin](#) - [WebConfig](#) - [Manual](#) - [Handbuch](#)



Archivista WebClient
Version 2009/1 - Powered by Archivista GmbH

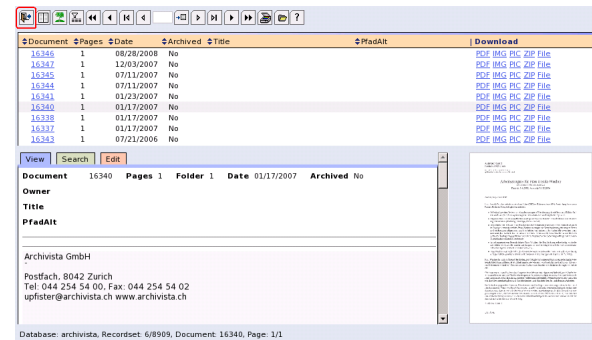
Host: localhost
Database: archivista
Username: Admin
Password: archivista
Language: English

To login please enter the password and click 'Login'.

➡ **Important:** The preset password is 'archivista' in lowercase letters.

For the functionality of the WebClient we refer to chapter 6.

5.2.3 Quitting the WebClient



You can quit the WebClient by clicking the first icon in the toolbar. This will take you back to the login screen.

5.2.4 Login to WebAdmin

ArchivistaBox is delivered with database 'archivista' containing already a set of fields. Whenever you would like to edit the structure of a database or to add or delete users, you must login to WebAdmin. Make sure that you are not logged in to WebClient (see above). Click the link 'WebAdmin' at the top left of the window.

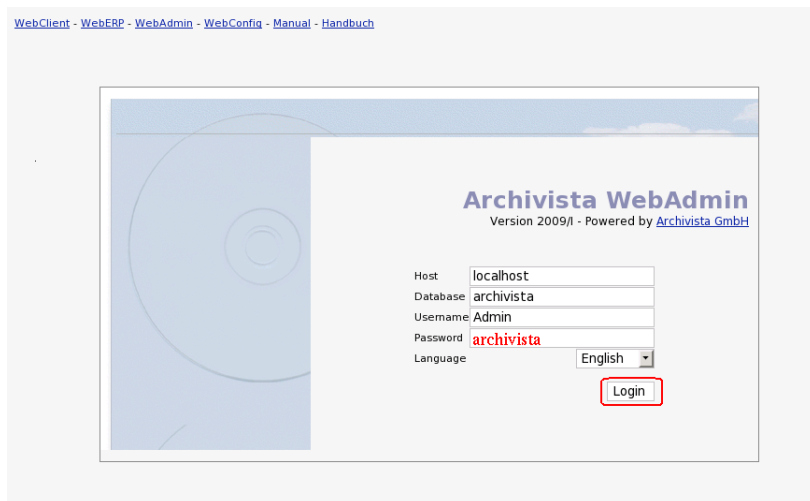
[WebClient](#) - [WebAdmin](#) - [WebConfig](#) - [Manual](#) - [Handbuch](#)



Archivista WebAdmin
Version 2009/1 - Powered by Archivista GmbH

Host: localhost
Datenbank: archivista
Username: Admin
Passwort: archivista
Sprache: Deutsch

You will see in front of you the login dialog of the application WebAdmin:

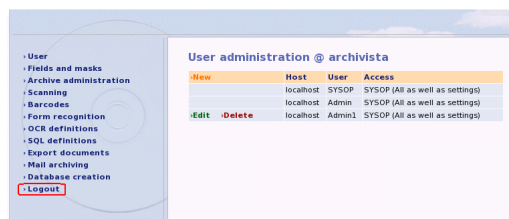


Enter the password 'archivista'. Subsequently, click the 'Login' button.

The functionality of WebAdmin is described in section 10 of this manual.

5.2.5 Quitting WebAdmin

To quit the application WebAdmin click 'Quit program'. You will be taken back to the login screen.

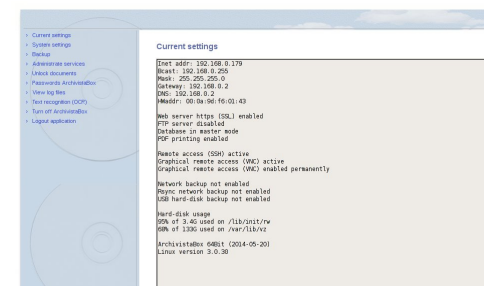


5.2.6 Login to WebConfig

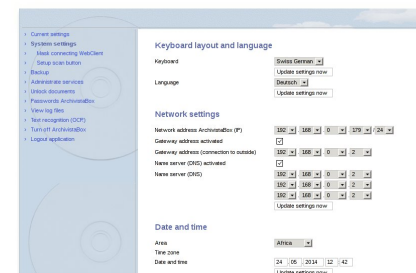
To connect ArchivistaBox to the outside world (network) you must start the application 'WebConfig':



In the login dialog enter the root password. Click the 'Login' button. You will see the main screen in front of you:

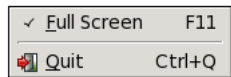


Here you should click 'System settings' in order to be able to select the language or add an additional DNS server.



5.2.7 Desktop

The system menu can be accessed directly from the screen of ArchivistaBox. You will either see a grey screen or the login dialog of the WebClient. If the login dialog of the WebClient is visible, you can close it by means of the right mouse key (or click on 'Quit').



There will always be a grey screen. One more click enables you to access the system menu.



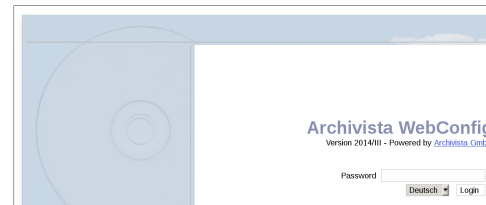
➡ **Difference WebConfig and system menu:** The application WebConfig (see 24 contains almost all menu items of the system menu. That is why you will need the system menu only rarely. Some points, however, can be reached only via the system menu. (E.g. restoring of backups).

More information on ArchivistaBox Desktop can be found at 25.

5.2.7.1 Accessing the desktop via WebConfig login

When the login dialog of WebConfig is invoked, there is a check if a) either the graphic remote maintenance (VNC) is active, if b) ArchivistaBox Dolder is active and/or if c) somebody is working locally. If this is the case, the link 'Desktop' will appear at the top between 'WebConfig' and 'Manual'.

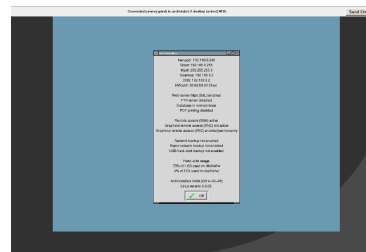
WebClient • WebAdmin • WebConfig • **Desktop** • Manual • Handbuch



After clicking on 'Desktop' you are requested to enter a password. Either enter the password for the graphic remote maintenance (VNC) or the Archivista password (in the case of ArchivistaBox Dolder).



After successfully entering the password you will be taken to the corresponding ArchivistaBox desktop.



As the entire process takes place in the browser, there are no plugins or auxiliary programs needed anymore to get to the desktop of ArchivistaBox.

5.3 Updating ArchivistaBox

ArchivistaBox is not installed or upgraded in the classical sense. ArchivistaBox is started from a file (ISO). If a customer would like to use new functions (release 2014/III), s/he lets the distribution

partner or Archivista GmbH know via e-mail, telephone or postal mail. Archivista GmbH will create the updated ISO file for the customer.

In the case of **ArchivistaDMS** it is necessary to log into **Web-Config**. Then click the button '**Online update**' under menu item '**Quit ArchivistaBox**'. The new release is fetched from the Archivista download server and installed so that the following start-up launches the new release.

If **ArchivistaVM** is used, the **online update** can be triggered from '**Server**' and subitem '**Update server**'.

➡ **In both cases status messages will appear after the process has been triggered. When these messages do not appear anymore, ArchivistaBox can be restarted.**

5.4 Notes on the box models

The ArchivistaBox models are available in different performance classes, for one or several users, for small as well as large archives. In the following we would like to present the different models and box types.

➡ **Important:** On the basis of today's fast paced developments in the field of hardware we reserve the right to change products. An updated list of the box models and types can be found in our shop: shop.archivista.ch.

All ArchivistaBox models contain the following components: text recognition and creation of searchable PDF files, file upload via FTP server, and WebClient (without access from third-party applications via API), versioning, print server (incl. automated takeover of file name). Furthermore, the module virtualisation (ArchivistaVM) is at the disposal of users.

All ArchivistaBox systems (except Dolder) can be extended with scan station Albis. The models Eiger and Matterhorn also allow extension with scan stations Mythen and Rothorn.

In addition, there are the following modules: barcode recognition, API: access via third-party applications, COLD=extraction of field values during the virtual printing process, mail archiving with IMAP protocol and form recognition. You will find these modules as options to be purchased with the ArchivistaBox models. See shop.archivista.ch.

5.5 Virtualised ArchivistaBoxes

Each ArchivistaBox (except the Albis, Mythen, Rothorn and Matterhorn ScanBoxes) can be purchased in a virtual form (without hardware). In such cases the key data (performance) of the corresponding model apply. With regard to the virtualised ArchivistaBox systems you can choose whether you want to run the virtualisation platform yourself (e.g. VMware, VirtualBox, etc.). In this case, you must ensure the operation of the virtualisation platform yourself. Alternatively, you can operate most ArchivistaBox systems using the ArchivistaVM module. In this case the actual ArchivistaBox (DMS) is operated in a virtual form within the ArchivistaVM module. For more information about ArchivistaVM see 26.

➡ **Important:** Please note that there are support restrictions for virtualised ArchivistaBox systems. For example, for virtualised solutions there is no support for directly connected scanners. These must instead be connected via scan boxes (see 32.3). Ideally, you should use ArchivistaBox Albis for this (see 5.13).

5.6 ArchivistaBox DOLDER

Hardware: DualCore processor, 2 GB of memory (RAM), USB, 3xLAN, hard drive with 30 GByte.



Software: Single-user version for a maximum of 20,000 files or 100,000 pages and one database (max. 20 GB of data). With an option to extend to two users and two databases.

➡ **Note:** ArchivistaBox Dolder possesses the same range as the other ArchivistaBox models but cannot be extended with modules. The maximum size of a file is currently 128 MB while it is 512 MB with the other ArchivistaBoxes.

5.6.1 Initial operation

Please note that ArchivistaBox Dolder does not have a display screen port.

After start-up ArchivistaBox Dolder can be accessed via the IP address that was indicated when the box was ordered. For this to work **the network cable must be connected to the port on the right hand side of the box.**



➡ The ports in the middle and on the left side (next to the USB ports) are not to be used. If the need arises, these two ports can later be utilised for further IP addresses (e.g. for data backup on NAS).

Note: Even if ArchivistaBox Dolder does not have a physical port for a monitor, it still has a virtual monitor. You can access it by going to the login screen of WebConfig. There you will find a link named 'Desktop'. By clicking this link you will be taken to another login screen (default password is 'archivista'). Subsequently, the virtual ArchivistaBox Desktop will appear directly in the browser, cf. 5.2.7.1.

Alternatively, you can reach the ArchivistaBox Dolder Desktop via the IP address with the TightVNC Viewer (optionally also with RDP) and the default password 'archivista' (as long as it has not been modified).

5.7 ArchivistaBox RIGI

Hardware: Dual-core processor, 4 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 160 GB.



Software: Two-user version for 20,000 documents and two databases (max. 20 GB of data). Total max. 500,000 pages. Expandable with the barcode identification module.

5.8 ArchivistaBox SAENTIS

Hardware: Dual-core processor, 4 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 160 GB.



Software: Multi-user version (standard scope 4 users, 10 profiles). 20,000 documents per database and four databases (20 GB of data). Total max. 500,000 pages. Optionally, other users/profiles and the barcode and API modules may be purchased.

5.9 ArchivistaBox PILATUS

Hardware: Quad-core processor, 8 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 160 GByte.



Software: Multi-user version (standard scope 4 users, 10 profiles). 200,000 documents per database and four databases (100 GB of data). Total max. 2 million pages. Optionally, other users/profiles and the barcode, API, mail archiving and COLD modules may be purchased.

5.10 ArchivistaBox TITLIS

ArchivistaBox TITLIS consists of two industrial-suited PCs forming together a redundant system. They each are equipped as follows.

Hardware: Dual-core processor, 8 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 320 GB.



Software: Multi-user version (standard scope 4 users, 10 profiles). 1 million documents per database and four databases (200 GB of data). Total max. 2 million pages. Optionally, other users/profiles and the barcode, API, mail archiving, COLD and forms recognition modules may be acquired.

5.11 ArchivistaBox EIGER

ArchivistaBox EIGER consists of two industrial-suited PCs forming together a redundant system. They each are equipped as follows.

Hardware: Quad-core processor, 8 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 750 GB.



Software: Multi-user version (standard scope 4 users, 10 profiles). 6 million documents per database as well as an unlimited number of databases (300 GB of data). Total max. 10 million pages. Optionally, additional users/profiles and the barcode, API, mail archiving, COLD and forms recognition modules may be purchased.

5.12 ArchivistaBox MATTERHORN

ArchivistaBox MATTERHORN consists of two industrial-suited PCs forming together a redundant system. They each are equipped as follows.

Hardware: Quad-core processor, 8 GB of memory (RAM), USB, LAN, VGA/HDMI, hard drive with 2 TB.



Software: Multi-user version (standard scope 4 users, 10 profiles). Unlimited number of files per database and unlimited number of databases (2 TB of data). Optionally, additional users/profiles and the bar code, API, mail archiving, COLD and forms recognition modules can be purchased.

5.13 Scan station Albis (optional)

ScanBox Albis is an optional scan station to be combined with all models from ArchivistaBox Rigi upwards. This small-scale scanning solution distinguishes itself by high performance at minimum

power consumption (3 watt). The maximum daily volume is at 5,000 pages. The recommended volume is at 2,000 pages.

Hardware: ARM, 512 MB of memory (RAM), USB, LAN, DVI, 8 GB flash memory,



ArchivistaBox Albis is currently the only ArchivistaBox that does not run completely in RAM mode (512 MB are scarce, after all, for the entire GUI). For more information cf. 4.5. Nevertheless, it can be conveniently updated via online update (cf. 5.3).

5.14 Scan station MYTHEN (optional)

ArchivistaBox MYTHEN acts as a simple scan station. The daily volume is 10,000 pages, including text recognition (OCR). Optionally available with barcode recognition. **Hardware:** Quad-core processor, 8 GB of memory (RAM), USB, LAN, VGA/HDMI.



5.15 Scan station ROTHORN (optional)

ArchivistaBox ROTHORN acts as a scan station. The daily volume is 40,000 pages, including text recognition (OCR). Optionally available with barcode and form recognition. **Hardware:** six-core processor, 16 GB of memory (RAM), USB, LAN, VGA/HDMI.



Part III

WebClient

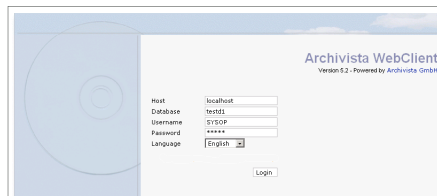
6 WebClient

Access to the Archivista database takes place via Internet browser.

6.1 Login

In principle, host name, database name, user ID and password must be entered. This information will be given to you by your system administrator.

➡ Note: It is possible that you only have to enter a user name and password. It is also possible that you may be able to enter the database name but not the name of the computer. Initially, all four values are required, although the first three have default values (localhost, archivista, Admin). The default password is 'archivista'.

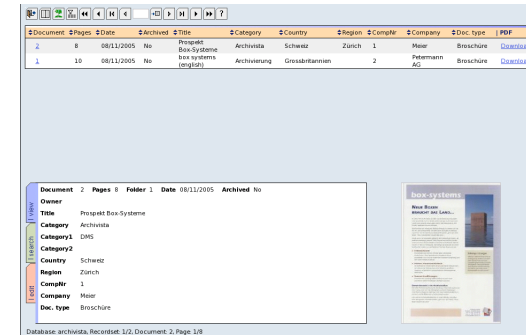


6.2 Views

There is a main view as well as a page view.

6.2.1 Main view

The main view consists of a list of documents (upper half of screen), details of the currently active document (lower half of screen on the left) and the image of the currently active page (bottom right).

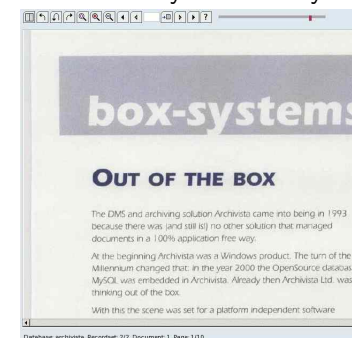


The status bar at the bottom left shows information on the currently active document: which document of how many it is, what number the document possesses and which page is displayed.

Depending on what rights you have, you cannot only search documents and pages but you can also edit and delete them. To the left of the document detail area you find tabs in different colors. To find out more information about searching please go to 8.2. To find out more about the editing mode see 8.3.

6.2.2 Page view

Page view shows you the currently active page. The image is automatically scaled to your browser window.



7 Navigation

Navigation, i.e. moving from one document to the next and 'leafing' through pages of individual documents, is not done by means of commands from the menu but by clicking the relevant icons.

7.1 Main view



Page view: display pages (F9). With this command you change to a view which shows the active page scaled to your browser.

↔ You can toggle between main and page view by clicking the image once.



Quit program (F12)



Select all records (F6)



Toggle between photo and table mode (Shift+F9). With this button the images can be displayed as photo album (instead of a table).



Toggle ratio table/page: With this button the table's width can be reduced to that of the detail view in order to obtain an enlarged view of the current page (image).



Browse backwards (Ctrl+F3). This command refers to the documents of the current selection.



View previous document (F3). This command refers to the documents of the current selection.



View first page (Ctrl+Shift+F3). This command refers to the pages of the currently active document.



View previous page (Shift+F3). This command refers to the pages of the currently active document.



Go to page (Alt+g). Before clicking the icon please enter a page number in the field in question.



View next page (Shift+F4). This command refers to the pages of the currently active document.



View last page (Ctrl+Shift+F4). This command refers to the pages of the currently active document.



Next record (F4). This command refers to the documents of the current selection.



Browse forward (Ctrl+F4). This command refers to the documents of the current selection. With this icon you can scroll forward through the list of documents.

7.2 Sorting

◆ Date
01/16/2006
01/16/2006
01/16/2006

The document list can be sorted by clicking on the field name in the red bar (table title line).

7.3 Page view



Quit page view: back to main view (F9).

. Please note that you can also change between main and page view by clicking the image once.



Rotate graphic to the left (Shift+F7).



Rotate graphic to the right (Shift+F8).



Put graphic on its head (Alt+u).



View entire page (Ctrl+F2).



Zoom out (Alt+w)



Zoom in (Shift+F2).



Previous record (F3). This command refers to the documents of the current selection.



View previous page (Shift+F3). This command refers to the pages of the currently active document.



Go to page (Alt+g). Before clicking the icon please enter a page number in the field in question.



View next page (Shift+F4). This command refers to the pages of the currently active document



Next document (F4). This command refers to the documents of the current selection.



Along with the pages sticky notes can be saved. Areas can be hidden and/or remarks can be added to the pages. Further information can be found at 9.8

➡Note: The button 'sticky note' is visible only if the tab 'Edit' in the page view is active. The addition of sticky notes requires editing rights for the document in question.



Continuously adjustable zooming: by moving the little red bar the page display can be easily adjusted.

8 View, Search and Edit

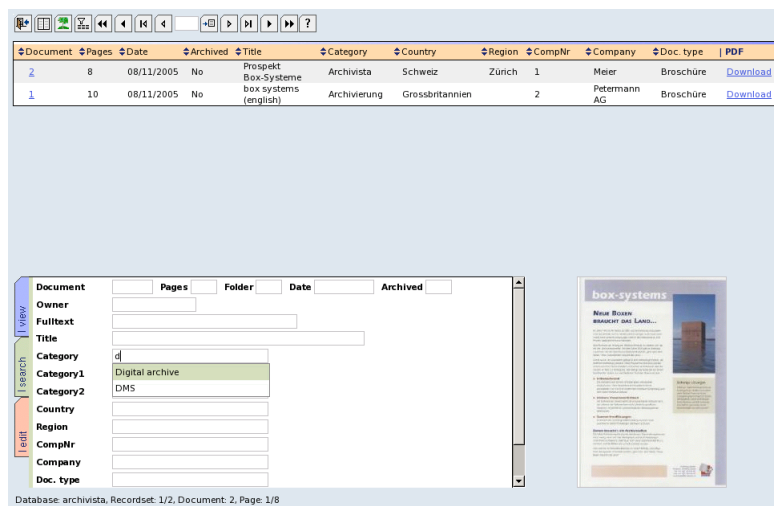
In addition to the main view and the page view, there are three further operating modes in the main view. Documents can be viewed, searched and edited. The currently active document is displayed on the lower right-hand side.

8.1 View (Ctrl+F5)

Click the blue 'View' tab in the detail view of the current document. (It is in the middle of the area on the left-hand side).

8.2 Search (F5)

Searching by means of the WebClient is very convenient: There is a special search dialog box at your disposal. Simply click the green tab of the detail view of the currently active document.



You see a dialog box with empty fields - ready for you to enter your search criteria. You can select documents according to your liking.

8.2.1 Full text search

If you do not want to search for your documents and pages on the basis of keywords in the fields but want to search the entire text of all pages, then make use of the full text search possibility.

Fulltext

The little window serving this special purpose is to be found in the upper right corner of the search dialog box.

Please note the following features of the full text search:

- There is no distinction between upper and lower case
- You can look for parts of words by adding an asterisk to a word; man*, for example, finds 'mandarin' as well as 'manager'
- An asterisk at the beginning of the search term is ignored; *man does not find 'Hopeman'
- A Boolean 'AND' can be achieved by first entering the one and then the other search term, e.g. word_writing. This search returns all pages on which 'word' and 'writing' are to be found
- You can enter expressions in between quotes; "Berlin_Wall" finds all pages in which exactly this expression appears.

➡ The full text search can only come up with results if the documents and pages are actually available in full text. If in doubt, please ask your system administrator.

8.2.2 Extending results or narrowing them down, new selection

8.2.2.1 Option 'extend'

Multilevel searches are possible by means of the options towards the bottom of the dialog box.

Type of query ☐ New selection ☒ Extend ☐ Narrow

When you do a second search and you tick the option **Extend** you will see the results of both the first and the second search. For instance: I click **New selection** and choose 'Sample_Inc.' in the field 'Client'. Then I click the **Search** button. After the results have been displayed I go back to the dialog box, I click Sample_Holding and in addition I click **Extend**. Then I click the **Search** button. This two-level search yields all documents that belong to Sample Inc. AND Sample Holding.

8.2.2.2 Option 'narrow'

Narrowing down my results works similarly. For instance: I click **New selection**, choose Sample_Inc. in the field 'Client' and press the **Search** button. After the results have come up I decide that I want to refine my search. I go back to the search dialog box and click **Narrow**. In the field 'document type' I choose `correspondence` and click the **Search** button. Now the system returns only the correspondence for Sample Inc. (All other document types like invoices, delivery notes, etc. are ignored.)

8.2.2.3 New selection

As mentioned above tick **New selection** whenever you want to search the entire document base for which you possess rights.

Type of query ☒ New selection ☐ Extend ☐ Narrow

As a rule, when you start up the program **all** documents appear that you are allowed to see. 'New selection' is the default value in the search dialog box. For workflow implementations it may make sense to display a selection that has already been narrowed down (only those documents are shown, for example, which need to be processed). In such a case please refer to the specific workflow documentation.

8.2.3 Date range

The format when searching for a date range is: d.m.yy-d.m.yy, for example 1.9.03-30.9.03. This stands for September 2003.

8.2.4 Entering ranges for number fields

You can also carry out a range search for number fields. This includes the Document, Pages and Folder fields. For example: By entering 10-11 in the 'Pages' field, you select all files that have 10 to 11 pages.

8.2.5 Searching with larger/smaller

When you carry out searches in date and number fields, you can also use < for smaller or equal to and > for larger or equal to.

8.2.6 Searching text fields for unequal

By including '!' at the start of your entry, you can search text fields for different files. For example, if you enter '!Meier' for

a company name, the system finds all files that are unequal to 'Meier'.

➡: If you work with wildcards, these are also used when searching for unequal fields. For example, if you enter '!tree', all hits that contain 'pear tree', 'tree trunk' or 'tree' are excluded.

8.2.7 Joker

At the very bottom of the search dialog box you can specify whether search strings entered in text fields should begin or end with a wildcard or joker symbol.

☒ Joker at the begin of textfields ☒ Joker at the end

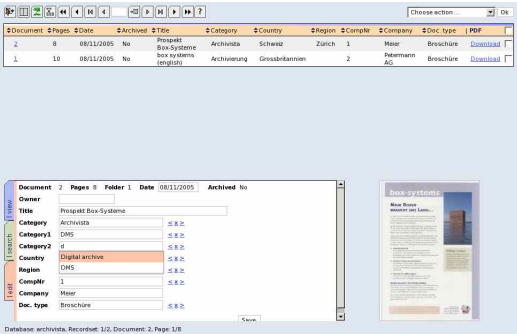
The use of this function is best explained by examples. Hence, when the option **Joker at the beginning of textfields** is ticked, then the search term `tree` returns also documents with 'Portree' or 'Crabtree' in the field in question.

When the option **Joker at the end of text fields** is activated, then the search term `tree` returns also documents with words such as 'Treetop', 'Treehouse' or 'Trees'.

➡ Please note that this option does not apply to the full text search field. For the variations possible in the full text search please refer to 8.2.1.

8.3 Edit mode (Shift+F5)

There is a special dialog box in the WebClient for the 'edit mode'. Click the red tab in the detail view of the currently active document. (You find it in the lower half and on the very left of your screen.)



You see the edit dialog box specially adapted to your and your company's needs. Depending on user access rights and work sphere some fields may be locked for you while others require filling-in with keywords. In addition, there are various ways in which to fill in fields. Below we give an overview.

To ensure stringent classification of documents the system guides and supports the individual user in the process of keyword allocation. The more data in a database the more important the streamlining of possible entries. Ask your system administrator to explain to you how your specific edit dialog box works.

In the following we give a quick overview of what kinds of entry possibilities the system typically provides. Should you want to know more we refer you to chapter 13 and 13.3.

8.3.1 Uncontrolled keywording

You are free to enter anything you like.

8.3.2 Field types: number field and date field

The field requires a number (no text) or numbers in date format (mostly mm/dd/yyyy or dd.mm.yy).

8.3.3 Controlled vocabulary

You need must choose from a given number of entries. As soon as you make an entry, a drop down list appears from which you can pick the value you want.

Category2

d
Digital archive
DMS

Users who are allowed to add to the pre-defined entries see the following characters next to the field.

< x >

By clicking > on the right hand side **new expressions** can be added. A field opens up.

<input type="text"/>	< x >	Project Wireless Lan	Add
----------------------	-------	----------------------	-----

Enter the new value, click the button 'Add' and the new expression is saved. If you want to stop adding new expressions, click < and the field closes again.

To delete controlled vocabulary, select the expression you have in mind and then delete it by clicking X.

8.3.4 Field type 'hierarchical'

Hierarchical fields allow you to create keyword trees - similar to the 'folder within the folder' principle of the Windows Explorer. Typically, hierarchical fields appear in the guise of pull down menus with controlled vocabulary.

8.3.5 Field type 'Multi'

The field type 'Multi' enables you to add multiple keywords per category. You can allocate one document to several customers,

for example. Multi fields typically appear together with controlled vocabulary.

8.3.6 Field type 'linked field'

For manual entry a linked field makes sense because unambigu-ousness may often only be attained if a number field is linked to a text field.

This field type occurs often together with barcode processing: a text field is linked to a number field and when the latter is filled in after the barcode has been 'read' by the system, the former can be completed accordingly. But also for manual entries a linked field can make sense because unambiguousness is often only achieved with the help of numerals.

➡ Example: a recruitment company possesses several hundred customers of which some have very similar names. 'Smith Vehicle Construction Ltd.' und 'Smith Vehicle Parts Ltd. ' can be distinguished more easily when the client number appears alongside the client name.

Also this field type occurs together with controlled vocabulary. When an expression from a text field is selected the value in the number field jumps automatically to the number that is linked to the expression. And when a number is selected, the text in the text field changes to the expression that is tied to the number.

Smith
07

When a new expression is added, clicking the X makes two fields open up:

Smith	< x >	Smith Vehicle
07	< x >	3849
Add		

To add an entry (or delete one) you must pay attention to both fields.

➡ Beside the combination 'text field'-'number code' there is also the combination 'text field'-'text code' for those cases where specific abbreviations are used in a company but where the abbreviations are not known well enough to exclude confusion. The combination 'Reporting and Consolidation'-'RC' elegantly enables correct keywording, for example. For the rest, the procedure is the same as with 'text field'-'number code'.

8.3.7 Navigation

To hop fast and easily from one field to the next in the edit dialog box you best use the tab key with your left hand.

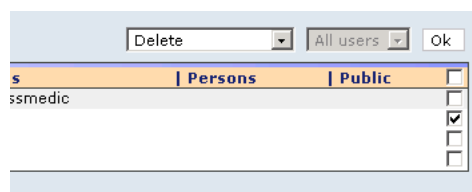
8.3.8 Document scanning

This is a special case of the Editing function insofar as it is required that a scanner be attached to the ArchivistaBox. This topic is treated in a separate section. See 9.4.

8.3.9 Deleting

While documents are edited in the edit dialog box, the 'delete function' is run from field 'Choose action...' in the upper right corner of the screen.

First go to the Edit mode by clicking on the red tab. Thereon the pull-down field 'Choose action...' will appear in the upper red corner. Select 'Delete' and click the little box(es) beside the document(s) you want to erase.



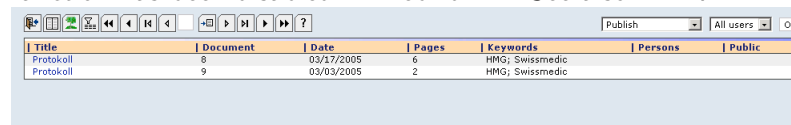
Click OK.

A dialog box will pop up to ask you whether you want to go ahead with the deleting process. Click OK again and deletion takes effect.

8.3.10 Publish and Unpublish

This function enables an owner of a document to make it available to an extended circle of users but without losing ownership of the document him- or herself.

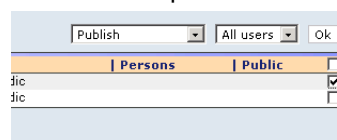
➡ If you do not find this menu item in the WebClient, then the function has been disabled in WebAdmin. See also 14.1.7.



In our example user 'freuler' clicks the red edit mode tab. In the upper right corner of the screen there appears a field with different actions. To publish documents to further users the action 'Publish' must be selected.

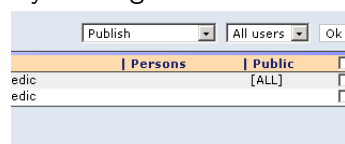
In a second step the user(s) for which the documents should be published are specified in the field adjacent to 'Actions'.

In a third step the document(s) in question are ticked.



➡ By ticking the small box at the top all documents of the current selection can be selected for the action that follows.

By clicking the button OK the action is carried out.

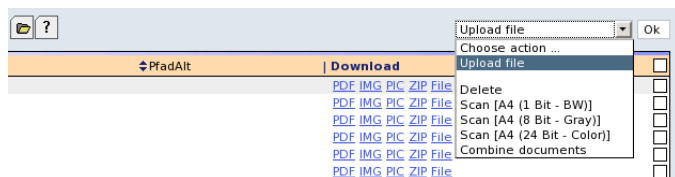


In our example one document was released for all users. This can be seen from the entry [ALL] next to the record concerned.

To reverse the publishing process choose **Unpublish** instead of **Publish**.

9 Extended functions

9.1 Uploading a file



You use this menu item (or optionally the icon shown on the left) to directly upload a file to the active database within the WebClient. After the call, the toolbar is displayed in a slightly different way.



You can use the first button to determine the file that you want to send to the archive. To the right of this, you can choose whether the file is to be processed with 'Black/White', 'Gray' or 'Color'. On the far right, you determine whether OCR text recognition is to be carried out and if so, what type.

🔄: You can upload PDF files as well as well-known bitmap files (tif, png, gif and jpg). As of version 2009/1, you can also upload all Office files. When you do this, these files are scanned automatically. The original file is also archived and you can check it out again later.

9.2 Copy template

Copy the fields of the **current document** with the help of **key F7**, navigate to a different document, change to tab **Edit** and paste the copied fields to the **now active document** by means of **key F8**. Please note that no new document is created but the active document receives new keywords. This action can be repeated at will in order to add keywords efficiently.

Alternatively, you can achieve the same by clicking on 'Copy template'. A new document will be created: It shows the same field information as the old document (template). Select the document (by ticking the check box on the far right): You can now start scanning into this document.

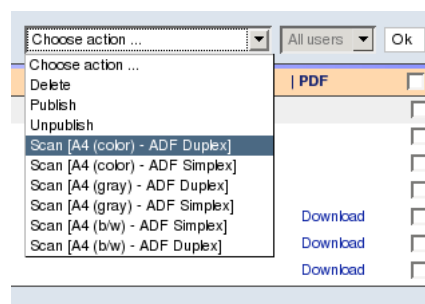
9.3 Create PDF file

You would like to create a PDF file from several existing documents. Select the documents in question and go to function 'Create PDF file' in the editing mode. After a certain time (the PDF file must be put together) you will see one PDF file in which all checked documents have been incorporated.

9.4 Creating and scanning documents

It is the Edit mode that enables you to create documents and to scan pages.

First go to the Edit mode by clicking on the red tab. Thereon the pull-down field 'Choose action...' will appear in the upper red corner. Select 'Scan'. Insert the document you want to scan into your scanner and click the 'OK' button.



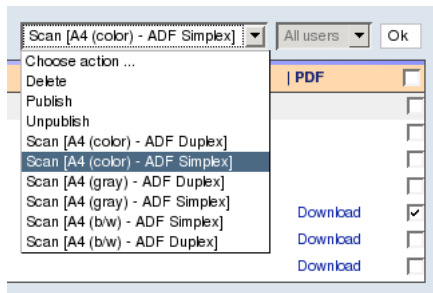
After about 30 seconds the newly created document and the scanned pages will appear. Click



or press F6 to refresh the screen and to see whether the pages are already there.

To add new pages to an **already existing** document follow the steps below:

- go to the Edit mode by clicking on the red tab
- select 'Scan' in the pull-down menu 'Choose action...' in the upper right corner of the screen
- tick the document to which you want to add pages



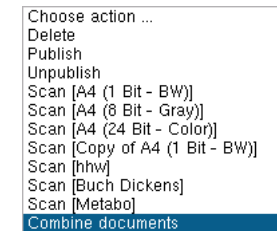
- click 'OK'
- confirm the pop-up message.

🔔: Scanning directly in the WebClient works only if a scanner is connected to the ArchivistaBox computer. If this is the case, a key pad will most likely be available for this ArchivistaBox, too. This allows for even more efficient scanning, see 9.17.2.

9.5 Combine documents

To add a new page to an existing document one can normally go to the document in question and scan the new page. With a digital copier this does not work, however. A separate document is created for the new page that has been scanned. To move this page to the target document you need the function **Combine documents**.

This function can be found in the field **Choose action** appearing in the top right corner when you go to the Edit mode.



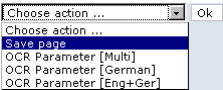
🔔: If you want to combine two documents, none of the two documents may have the status 'archived'.

To combine two documents you must proceed in the following manner:

1. Go to Edit mode
2. Tick the two documents that you intend to combine. The document to which the other one will be copied is underlaid with gray. You can change the target document by doubleclicking the target document.
3. Go to the **Choose action** field, select **Combine documents** and click Ok

9.6 Editing and deleting pages

You can edit the pages of a document: It is possible to rotate and save individual pages, to delete them or to assign specific OCR definitions to them. All of these functions are carried out by means of the 'Choose action' menu in Page view.



9.7 OCR Parameters



When pages are added, each page carries a flag that determines later on which options apply when the function 'text recognition' is run. The options can be defined in WebAdmin under 'OCR Parameters'. See 17. For each OCR definition there is a menu item with which the desired definition can be allocated to the current page.

By means of the menu item **OCR Parameter [use for document]** the selected OCR definition can subsequently be allocated to the entire document. This is practical when during the scanning process a wrong OCR definition was active (for example English instead of German for documents in German). In such a case the text recognition engine will not recognise umlauts in the first run. With these OCR Parameter options things can be put right: In a second run the umlaut letters will be correctly interpreted. The menu item **OCR Parameter [use for selection]** will transfer a new OCR definition to the entire current selection of documents.

➡ **Important:** After allocating an OCR definition to pages or documents, the text recognition engine has to be started once more. This can be done in WebConfig at 24.9.

9.8 Notes

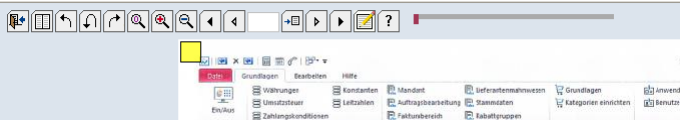
'Sticky notes' enable the user to add information directly to a document page. This works for adding textual information but it can also mean hiding confidential data with rectangles or highlighting something specific.

These notes are managed in 'Page View' via the button below.

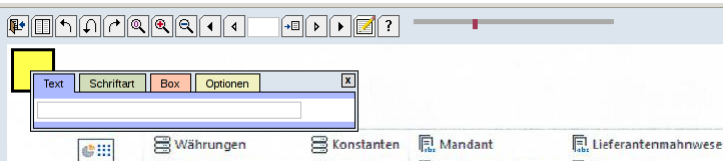


➡ **Important:** The button above is visible only if previously the tab 'Edit' has been activated in the 'Main View'. In 'Page View' the button can be activated with function key 'F5'.

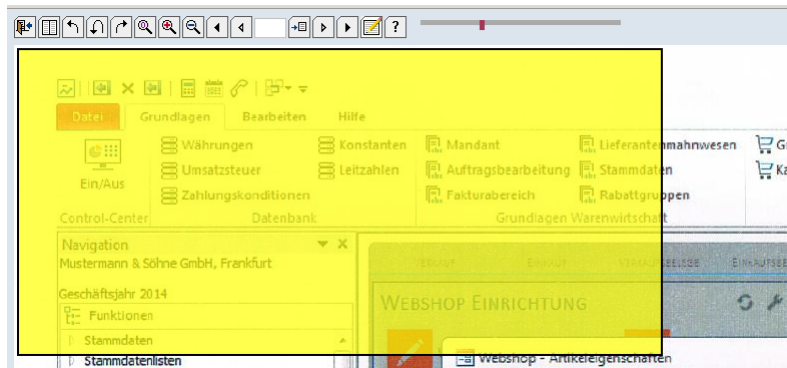
➡ Normally, information on a page in a database should be captured in a structured way by means of fields, not by unstructured comments. However, in certain cases (for example for print-outs or for forwarding a digital document) sticky notes may prove useful. To create a note click on the button. Subsequently, a yellow rectangle will appear on the upper left hand side of the page.



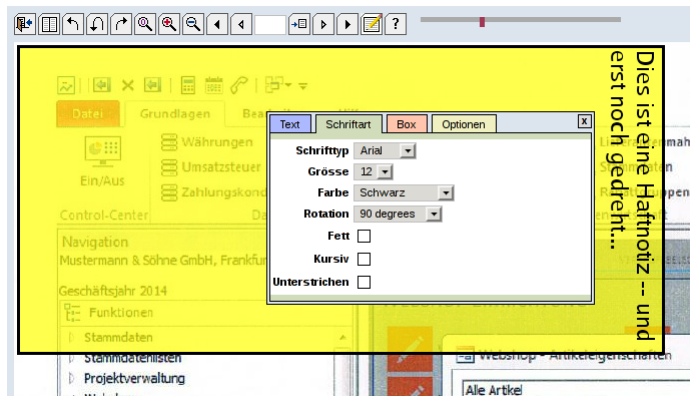
To edit the note click into it with the right mouse key. The option window will appear.



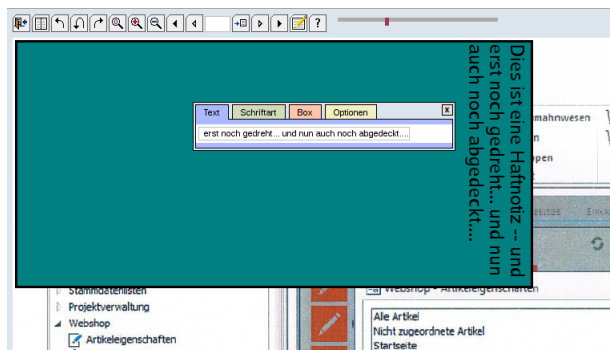
The size of the note can be modified by means of the mouse.



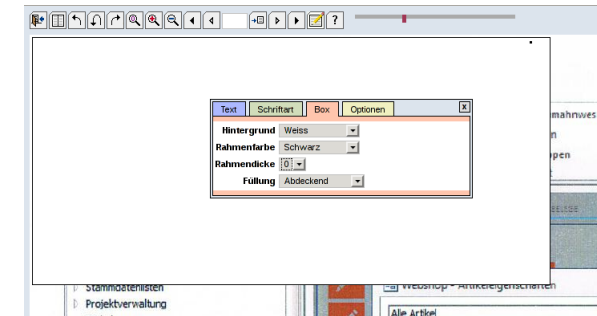
It goes without saying that text can be added. Optionally, it can be rotated.



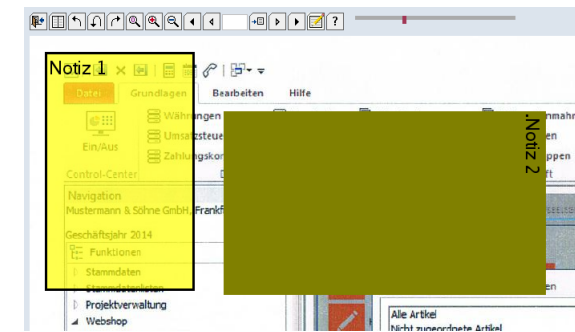
It is further possible to completely hide the background.



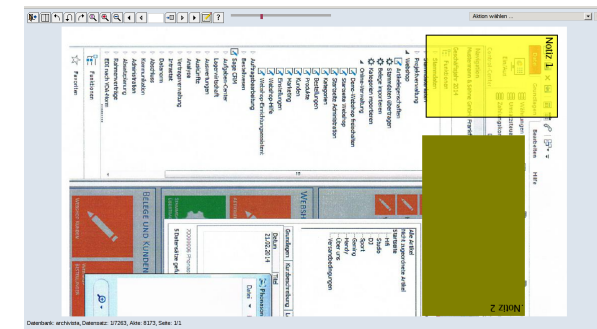
Often, specific areas need to be covered. Below you see an example where an area has been covered in white (black would work too, of course).



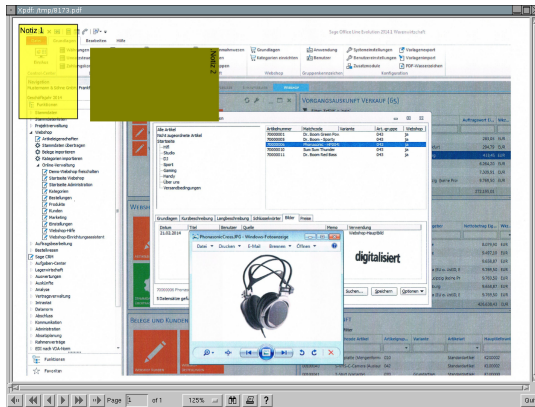
To a single page a maximum of 32 notes can be added. They can have completely different attributes.



Once created, notes are anchored. They zoom with the page and when the page is rotated, the notes follow the rotation.



Notes can always be modified, i.e. they are not written into the image. Solely, when a PDF file is created, the notes are merged with the page image, the idea being that a page area hidden underneath a black rectangle should definitively not allow anybody to decipher anything anymore.



9.9 Text recognition for page (with notes)

Sometimes text recognition is not successful straight away because automated identification of a text block may go wrong and the result is less than satisfactory. A text block inside an illustration or an image can 'get lost' because the entire graphic representation is recognised as an image.

In such cases text recognition can be triggered manually by highlighting the text blocks with notes. Subsequently, one can run menu item 'Text recognition for page (with notes)' and the recognition is carried out for the highlighted areas only. This menu item can be found in the actions menu in Page View. It is visible when tab 'Edit' is active in the Main View.

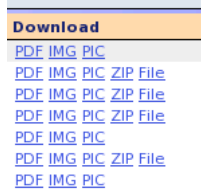


The illustration above shows several highlighted text blocks intended for text recognition. After the command is run, the text blocks are processed. The extracted information appears promptly in tab 'View' – and becomes available for the search functionality.



Afer text recognition the notes should be removed. Otherwise text will be obliterated in PDF files – in analogy to what was said further above about sticky notes.

9.10 Download



There are a few links on the right-hand side of the table in the main view of the WebClient. You can use these to obtain the respective document or a specific page as a PDF file or as an image file.

At present, you can download the following formats:

- PDF: Page/file as a PDF file. Depending on the option 'Convert whole file to a PDF file', the system provides the current page or the whole file as a PDF file.
- IMG: The current page is downloaded in high resolution.
- PIC: The current page is downloaded in reduced resolution.
- ZIP: An original file (source) that can be downloaded as a zipped file exists for the respective file.
- Mail: The current file was archived as a mail message. You can use this command to return the file to the mail server.
- File: The current file was archived as an Office file. By clicking on the link, you can restore the original file in the hard drive or open it directly.

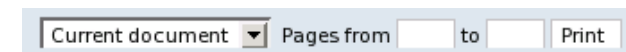
➡ If the 'Create PDF files' option is activated in WebAdmin, the PDF file can be downloaded directly. If this is not the case, the respective PDF file is created 'on the fly'.

9.11 Printing pages

In principle you can print individual pages by going to 'Page view', displaying the page in adequate size and running the print function from the web browser menu. For individual pages this is workable, however, if you need to print one or several documents in the WebClient, then it is inconvenient.

➡ This option must be activated by the administrator, see 11.1.13.

If the administrator has activated the option and thus enabled the printing function for you, you see at the bottom right of your screen a few fields, which you can use to print one or several documents.



➡ You can print more than one document only if you possess 'Editing' rights. If you do, you can select the documents in question and run the option 'Selected document(s)' under 'Current document' to print them all at once.

9.12 Barcode processing

If your Archivista WebClient 2014/IX has been equipped with the barcode processing module, then you can easily scan your pages and keywords are added automatically. See section 16 for more information about barcode processing.

Simply insert the documents with the barcodes on them into the ADF of the scanner and scan them. Subsequently, the system will read the barcodes, decode them and write the data into the fields in question. More information on this topic you will find under 16.

9.13 Versioning

ArchivistaBox gives you the possibility to modify previously uploaded files and upload them again. More on this can be found in section 9.10.

For ArchivistaBox 2010/IV a new versioning function is available that works with all Office and PDF files. Versioning means that documents that have been uploaded, modified later on and re-uploaded are automatically provided with a version number. This way you can keep track of the entire history of a document (all versions) in the WebClient.

➡ Versioning must firstly be turned on. To do so, see 14.4.

Once you have accessed a file with active versioning from the WebClient, you will see the file not only with the file name, but preceded by a file stamp. In the above example, the name is 'eins.pdf'. When you first check out, you will receive the file name 'Doc_Version.Name.pdf'. You could now edit this file and then upload it again. Now note the view below.

The screenshot shows the ArchivistaBox WebClient interface. At the top, there is a navigation bar with icons for home, search, and other functions. Below this is a table listing documents. The table has columns for Document, Pages, Date, Archiv, Title, Versioningkey, Versioning, Filename, and Download. The document 5488 is highlighted, showing it has 1 page, a date of 03/04/2011, and is titled 'eins.pdf'. Below the table, there is a detailed view of the selected document. This view shows the document's title, versioning key, and filename. It also displays the document's content, which is a PDF file named 'eins.pdf'. The document is dated 18th January 2008 and was created by Archivista GmbH, CH-8118 Pfaffhausen. The web pages are www.archivista.ch.

Document	Pages	Date	Archiv	Title	Versioningkey	Versioning	Filename	Download
5488	1	03/04/2011	No		1.1	5487	eins.pdf	PDF IAG PC ZIP File
5487	1	03/04/2011	No		1.0	5487	eins.pdf	PDF IAG PC ZIP File
5486	1	02/26/2011	No					PDF IAG PC
5484	1	02/18/2011	No					PDF IAG PC
5483	1	02/18/2011	No					PDF IAG PC
5482	1	02/18/2011	No					PDF IAG PC
5481	1	02/18/2011	No					PDF IAG PC
5480	1	02/18/2011	No					PDF IAG PC
5479	1	02/18/2011	No					PDF IAG PC
5478	1	02/18/2011	No					PDF IAG PC
5477	1	02/18/2011	No					PDF IAG PC
5476	1	02/18/2011	No					PDF IAG PC

Document 5488 Pages 1 Folder 3 Date 03/04/2011 Archived No

Title

Versioningkey 1.1

Versioning 5487

Filename eins.pdf

Manual

Archivista 2008/8

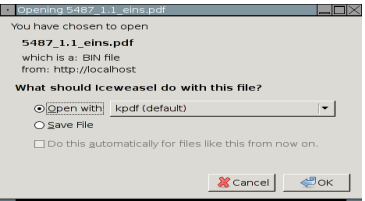
© 18th January 2008 by Archivista GmbH, CH-8118 Pfaffhausen
Web pages: www.archivista.ch

Database: archivista, Recordset: 1/718, Document: 5488, Page: 1/1

You can now see the file 'eins.pdf' twice in the archive, once with version '1.0' and also with version '1.1', i.e. the file was imported as a new version when you uploaded it again. As stated, this process runs entirely in the background. Note the field 'Title'. This field contained the value '450' in version 1.0. For the later

version that has been uploaded, the field values of the previous version are automatically copied to the new version, i.e. the file does not need to have keywords re-assigned. Unless you expressly wish to assign different keywords to the new file.

You can, of course, check the files in and out several times. To do so, click on version 1.1; you will then receive the file as follows:



You should bear in mind that you can only re-import one version at a time and that you should always work with the latest version. If you export and re-import an older version, that version will naturally be processed, but the version number remains the old version, with a minus sign added (e.g. version 1.0 would become version -1.0).

9.14 Attributing keywords with F7 and F8

Thanks to the new WebClient of ArchivistaBox 2012/IX fields can be copied from any document by means of function key 'F7'. Subsequently, the field values can be pasted to any other document in edit mode by means of 'F8'. Searching for documents is simpler, too.

9.15 Search terms of previous search

The last search is stored in the search dialog and can be rerun (also in modified form).

9.16 Replace

Now and then you will want to replace field values globally. The WebClient offers the dialog 'Replace'. Any field values can thus be replaced at will for the current selection of documents – in a way that keeps data consistency intact. This means (to name only one example) that when a linked field (e.g. CompanyName and CompanyNumber) is modified, always both fields are kept up to date. This feature requires administrator's rights (SYSOP).



9.17 Scanning documents

In the WebClient scan definitions are available in the 'Choose action' drop down list. You must be in edit mode. If the scanner is connected to the primary ArchivistaBox (where the documents are stored), this is an easy way to scan documents.

9.17.1 Scanning with scan station

In the case of several scan stations the scanning process must be triggered from the specified scan station. This means that firstly, the IP address of the scan station must be entered in the web browser. A user name is visible in the WebClient (most often the user 'scan' (password is customer specific)). After login, the scanning process can be started via 'Edit' and the actions can be run. The scanning process is triggered and the documents are automatically transferred from the scan station to the main box.

9.17.2 Scanning with KeyPad (keyboard)

.

Sometimes one would like to scan something but the WebClient is not open. Similarly, it may seem (too) complex to select the correct IP address when there are many scan stations. In such cases one can utilise a numeric keypad (keyboard works too): By entering a code one can trigger the scan process directly, a connection to the corresponding WebClient need not be available.

The following codes must be used (**afterwards 'Enter' must be clicked**):

- Number: corresponding scan definition of first database
- Number1.Number2: Number1 corresponds to the number of the database, Number2 to the desired scan definition
- Number*1: Number=definition, standard (*), feeder (1=simplex)
- Number*2: Number=definition, standard (*), feeder (duplex)
- Number*3: Number=definition, standard (*), feeder (landscape 270/90 degree)
- Number*9: Number=definition, standard (*), flatbed mode
- Number+1: Number=definition, color (+), feeder (1=simplex)
- Number+2: Number=definition, color (+), feeder (duplex)
- Number+3: Number=definition, color (+), feeder (landscape 270/90 degree)
- Number+9: Number=definition, color (+), flatbed mode
- Number-1: Number=definition, b/w (-), feeder (1=simplex)

- Number-2: Number=definition, b/w (-), feeder (duplex)
- Number-3: Number=definition, b/w (-), feeder (landscape 270/90 degree)
- Number-9: Number=definition, b/w (-), flatbed mode

In the case of the first database the database number (x.) need not be specified but it can be specified. Even if one scans to another database (x.y+2) the other options color and feeder can be used. Furthermore, one can add a slash /. The effect of this is that an individual document is created for each sheet. Example: 2.1-2/ results in scanning into the second database with the first scan definition, in b/w and in duplex mode. The final slash makes the program create individual documents for each scanned sheet (equalling two pages in the case of duplex).

➡ The number of the database is determined by the created databases (separated by commas) when the scan key is defined (see 25.14.1.8 and 24.3) under 'Database'. The string 'archivista,buha,private' means that with scan key 1.x or x scanning takes place into the database 'archivista', with 2.x into the database 'buha' and with 3.x into the database 'private' whereby x stands for the desired scan definition and the options described above.

➡ Normally, this function is activated for scan stations Albis, Mythen and Rothorn as well as ArchivistaBox Dolder; the scan key can in addition be turned on or off. See 25.14.1.8 and 24.3).

➡ **The cancel command:** Should you mistype a code (ideally being two- or three-digit), do not click the Enter key but click four times 9 (9999) or another non-sensical combination and then 'Enter'. This cancels the job because the scan process is triggered only if a valid code is supplied.

Part IV

WebAdmin

10 WebAdmin

Welcome to the administration of Archivista 2014/IX by means of the Archivista web interface! In the following we guide you through the different menus and options. To adapt Archivista 2014/IX to your needs you can set up users, create new fields and new (input and search) masks. Archivista 2014/IX is fully multi-client enabled, i.e. you can house several very different archives in one database.

To begin with, we describe how to go through the login and logout procedures of the WebClient administration.

10.1 Login



Archivista WebAdmin
Version 5.2 - Powered by Archivista GmbH

Host: localhost
Database: archivista
Username: Admin
Password: archivista
Language: English

Login

The database that needs setting up is accessed via the Internet browser. Host, database, user id and password must be entered. Please note that it is incumbent on you as systems administrator to keep track of all databases and sysadmin passwords.

➡ It may well be that in the framework of your particular installation host and database name need not be entered. The preset password is 'archivista' in lower case letters.

10.2 Logout



To logout you simply click **Logout** at the very bottom of the WebAdmin menu.

11 User

11.1 Administration (in general)

After clicking **User administration** in the menu you see a table of existing users.

User administration @ archivista

>New	Host	User
>Edit	localhost	SYSOP
>Edit >Delete	localhost	Admin
>Edit >Delete	192.168.0.%	heidi
>Edit >Delete	192.168.0.%	peter
>Edit >Delete	192.168.0.%	clara

The first cell in the upper left hand corner shows 'New' printed in red. Click 'New' to create a new user. You will see the following dialog box:

User administration @ testdb

Host	<input type="text" value="localhost"/>
User	<input type="text"/>
Password	<input type="text" value="New password (empty password allowed)"/>
Access	<input type="text" value="Read public ones, edit own records"/>
Create new documents	<input checked="" type="checkbox"/>
New documents with owner	<input type="text" value="All"/>
Group(s)	<input type="text"/>
Internal pages	<input type="checkbox"/>
Web	<input checked="" type="checkbox"/>
Workflow	<input type="checkbox"/>
No. of records following SQL queries	<input type="text" value="1000"/>
Mask definition	<input type="text" value="Admin1"/>
SQL definition	<input type="text"/>
E-mail account	<input type="text"/>
Add. field	<input type="text"/>
Add. notes	<input type="text"/>

The individual settings are described below.

11.1.1 Host

Here you enter the name of the host on which your Archivista database lies.

11.1.2 User

Here you enter the name of the user you want to create. The user name must be composed of letters but may also contain numbers.

11.1.3 Password

When a user is being created, the initial password can be set here. Similarly, the password can be reset here if necessary. If the field remains empty, the current password is not changed (in the case of a new account, an empty password is set).

11.1.4 Login

There are three options:

- Normal login
- New password (empty password not allowed)
- New password (empty password allowed)

When creating a user choose the second or third possibility. If you change settings later, then the first option makes sense. If a user forgets her password, you as SYSOP cannot access it but you set the second option above and the user is forced to enter a new password on next login.

11.1.5 Delete password

With this option a user's existing password can be reset.

11.1.6 Access

The possibilities are:

- Read public ones, edit none
- Read public ones, edit own records
- Read all, edit own records
- Read all, edit all
- Sysop (all as well as settings)

Each document possesses an owner. Normally the owner is assigned automatically: see 11.1.8. We can imagine many different archives:

1. For instance one with two users and both work exclusively with their own documents.
2. For instance one with two users none of which 'owns' any documents but where there are only public documents.
3. For instance one with three users. Let us imagine a florist, a bookkeeper and a general manager. The florist creates a photo collection for his own purposes, the bookkeeper scans documents posted to an account but puts items of common interest (like the list of public holidays for the new year) at the disposal of everyone and, finally, the general manager compiles records for her own specific uses but is certainly allowed to view all documents.

In the first case both users obtain the following access rights: 'Read public ones, edit own records'. There are no public documents.

In the second case both users obtain the access rights 'Read all, edit all'.

In the third case the users are given the following rights:

- Florist: 'Read public ones, edit own records'
- Bookkeeper: 'Read public ones, edit own records'
- General manager: 'Read all, edit own records'

This way the florist and the bookkeeper see only their respective (and the public) documents. The general manager can view all documents. When the bookkeeper creates a document she would like to make available to the florist, then she sets the owner of that document manually to 'ALL'.

11.1.7 Create new documents

If the user is allowed to create new documents, then this option must be ticked.

11.1.8 New documents with owner

This option is related to 11.1.6. The following alternatives are the most likely choices:

- New documents with 'own' owner
- New documents with owner 'ALL'
- New documents with group owner

11.1.9 Group(s)

Groups make those documents available to a user that are not really his own but belong to a larger entity. Typical groups are for example the accounts payable department or the production department. Let us take the example of Joe Smith who is part of accounts payable. Under 'New documents with owner' Joe Smith could have 'AP' and with access rights 'Read public ones, edit own records' Joe could handle all documents of the accounts payable department as if they were his own. And his colleagues can do likewise.

11.1.10 Mask definition

The option '**Mask definition**' serves to specify the input mask a user sees after starting up Archivista. By being able to select different input masks for different users you can have several archives in one database. (Multi-client capability)

11.1.11 SQL Definition

Here normally 'AVSTART' is chosen, i.e. you choose the definition which determines which document selection a user should see after starting up Archivista. The definition itself is made in a separate input form. See chapter 18.

11.1.12 E-mail account, additional field and additional notes

You can collect one e-mail address per user. The e-mail address may be used by third-party modules (e.g. workflow module) when certain events trigger user notification. 'Additional field' means that you can allocate an additional identification number to a user, if need be, and in the memo field 'Additional notes' you can enter any text you like.

11.1.13 Additional field

For those users who need to be able to make printouts the administrator can add a printer in the additional field. It must be pointed out that as of today only postscript printers are supported. The printer must be set up correctly as CUPS printer on the ArchivistaBox.

11.2 User administration (external)



There is a distinction between the external user administration and the internal user administration with MySQL. If you log in to your ArchivistaBox, for example, you directly access the MySQL database i.e. you work internally.

In all those cases where a central user administration tool exists (e.g. LDAP), ArchivistaBox can be connected to it. At every login the WebClient will check with the computer specified whether the account that is used exists and whether it is valid. Only if this is the case, login can take place.

At this stage, an internal user account will be created on the basis of information from the external server. Archivista specific fields of the user account will be copied from a template account that already exists in Archivista.

Use this menu function to define how you would like to handle your user administration.

11.2.1 Administration

When you click 'Administration' you will see the following:

Administration @ archivista

Access mode: MySQL

Server:

Port:

Program:

Domain for users:

Base DN for LDAP:

The first field '**Access with**' gives you the choice of the following options:

- MySQL
- LDAP
- HTTP query

MySQL

When you choose **MySQL**, access will always happen via the ArchivistaBox in question, i.e. no external computer will be used for login.

LDAP

LDAP (Lightweight Directory Access Protocol) is an application protocol that is widely used in computer technics. By means of the LDAP server it is possible to run user administration from one central system. Only one single password is necessary. The web client serves as a data transmission device between the central system and the individual database. Hence, LDAP allows you to administrate your users centrally.

Background information: LDAP allows the query and modification of information of a directory service by means of the TCP/IP network. One important characteristic is its hierarchical setup,

which for example MySQL does not use (MySQL possesses an ordinary, tabular structure). The data structure of an LDAP directory is like a hierarchical tree with roots, branches/twigs and leaves. The root (root, suffix) is the topmost data object, below it one finds the higher structures. If you allocate to a user two groups of different hierarchical status (see 11.2.2.1), then the hierarchical status of the user corresponds to the hierarchical status of the one of the two groups which is hierarchically highest.

HTTP query

Instead of the LDAP query described above, an external HTTP server (web server) is queried for user information. The query has two parts: password verification and user data query.

11.2.1.1 Server

Here you enter the IP address of the server, e.g.: 192.168.0.12

If HTTP is used, this field contains the complete URL of the web service that checks user ID and password, e.g. `http://my.server.com:1234/cgi-bin/pwcheck.pl`

11.2.1.2 Port (LDAP)

If needed, you can enter the interface (port address) here, e.g.: 389

Ignored for HTTP - the port can be specified in the URL (see above).

11.2.1.3 Program (HTTP)

If access should take place via a HTTP query, you have here the possibility to run a program that gives back a table with all users (name, email, group list).

11.2.1.4 Domain for users (LDAP)

Here you can specify from which branch of the LDAP tree the search must start. Enter the branch, e.g.: dc=archivista, dc=ch

11.2.1.5 Base domain (LDAP)

Specifying the domain is necessary in order that the LDAP server may know in which network it is, e.g.: archivista.ch.

11.2.1.6 Users small / Groups large

You can use this field to determine that when a user logs in with a mixture of upper and lower case letters, this is always converted to lower case. For example, the user is opened with 'meiere'. If this option is active, 'MeierE' is also accepted. No conversion is carried out if the whole name is entered in upper case.

11.2.1.7 User (Files without owners)

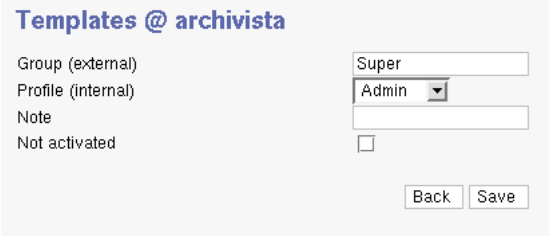
Here, you can explicitly specify a user name so that it is no longer possible to open a file in the ArchivistaBox and any services (scanning, FTP upload and PDF printer) that does not have an owner.

11.2.1.8 Setting a user for uploaded files

If this option is activated, the user that is logged in is set as the owner of the file when the file is uploaded.

11.2.2 Template

When you click on 'Template' you will see the following:



This screen lets you define from which internal Archivista user profile user settings should be copied. For the setup of an internal user account it is absolutely necessary that a matching template exists.

11.2.2.1 Group (external)

Please enter the name of a user group as it is supplied by the external server (LDAP or HTTP). If the external HTTP server returns a list of groups (e.g. 'SALES, MKTG, ACCOUNTING'), only the first group in the list is checked for a matching template.

11.2.2.2 Profile (internal)

Please select an account from the list that is to be used as a template for the new user account.

Note - if the account selected as the template contains data in the Group field, this is copied into the newly created user account; if it is empty, the group information supplied by the external server is used.

11.2.2.3 Remark

You can add information on the user in question by entering it in this field.

11.2.2.4 Not active

By clicking this check field the settings you have made remain inactive.

Do not forget to save your changes to the settings!!!

12 Field definition

After clicking **Fields** in the menu you see a table of existing fields. To make changes to an existing field click **Edit** or **Delete**. The first cell in the upper left hand corner shows 'New' printed in red. Click 'New' to create a new field. You see the following dialog box:

The image shows a web-based dialog box titled "Field definition @ testdb". It contains four input fields: "Field name" (a text box), "Field type" (a dropdown menu with "Text" selected), "Length" (a text box), and "Position after" (a dropdown menu with "Pages" selected). At the bottom of the dialog are two buttons: "Back" and "Save".

Individual settings are described below.

12.1 Field name

Please note that the name must be unique, must not contain any blank spaces and the first character must not be a number. The name of the field can be changed during the same session but after leaving the form it cannot be changed any more. Should it be absolutely necessary to change it later, it can be done from the MySQL console.

12.2 Field type

Decide for one of the following field types:

- Text (1-250 characters)

- Number integer
- Number double (meaning double-precision floating point to express fractions)
- Date
- YesNo

The field type cannot be modified in later sessions.

12.3 Length

When the field type is 'Text' you may vary the length of the text by entering a value here. Cannot be modified in later sessions.

12.4 Position after

Designates the position of the field within the table with respect to the other fields.

13 Mask definition

After clicking **Masks** you see a screen where you can select the mask which you want to manipulate further. The table below shows the fields of the active mask.

Mask definition @ archivista

Mask

	Field name	Label	Position	Width
New				
Edit	Persons		0	0
Edit Delete	Keywords		0	0

13.1 Mask

Here you can specify whether you want to

- select an existing mask and edit it
- delete an existing mask
- rename an existing mask
- create a totally new mask

Before you can make changes to an existing mask you have to select it. To rename a mask you select it first and then write the new name into the field in front of the button 'Rename'. Then, click 'Rename'.

Please note that you must choose a unique name which does not contain any blank spaces and does not begin with a number.

Whether you create a new mask or make changes to an existing one, there are always the following settings to consider.

13.2 Field name

You can add fields to the mask. Simply pick an existing field from the list.

13.3 Field type

Decide here about add-ons to field definitions. Does a field need to be linked to another field? Do you want to establish a set of fields that are hierarchically dependent on each other? The following sections describe the different types.

13.3.1 Normal

The default value is **Normal**. The field is displayed as a simple entry field.

13.3.2 Text code

You can create fields which are linked to other fields. One alternative is the linking of a text field with a text code field.

13.3.3 Definition

The field type '**Definition**' is used in a number of cases. The simplest is that of a combo box from which to select a predefined value.

The subsequent example elucidates the functionality. Go to 'Field definition' and create a new field named 'Department'. Specify the field type by selecting **Definition**. Now save the mask and logout from the WebAdmin tool.

On entering the WebClient you now see the following mask.

Document 28 Pages 0 Folder 1

Date 07/06/2005 Owner

Title

Department < x >

Save

Click the field 'Department': As there has not been defined any controlled vocabulary as yet, the pull down list is empty.

Click > and a field will open up. Now enter the value 'Purchasing':

Department < x > Purchasing

Then, click 'Add'. The value 'Purchasing' has been added.

Let us add the values 'Production' and 'Sales' to our field 'Department.' Now you should see a screen roughly like the following:

Department < x >

- Purchasing
- Production
- Sales

➡ Please note that when choosing a value from your drop down list it is sufficient to enter the beginning of the word up to the first distinctive character. The value you want is selected and by clicking 'Enter' you transfer it to the field.

13.3.4 1:N

The field type '1:N' enables you to build drop down lists which depend hierarchically on specific entries in other drop down lists, i.e. with this field type you can compile veritable keyword tree structures. Let us make an example. We go to the menu 'Field

definition' and create an additional field which is dependent on the field 'Department': let us call it 'Region.' For this field we activate **Field type '1:N'** and we choose 'Department' in the 'Link field'. We have thus specified a 1:n relationship between the fields 'Department' and 'Region.'

Field name

Field type

Link with field

Label

Position

Width

Save the changed mask definition. Go to the main view. Under 'Department' choose 'Purchasing.' Now create entries like 'France', 'Benelux' and 'Switzerland' in the field 'Region'. Finally, select 'Benelux'. Do not forget to save your entries.

Department < x >

Region < x >

- France
- Italy
- Germany
- Switzerland

Create a new document. Under 'Department' select 'Production' from the combo box. Now go to the field 'Region': You will notice that the controlled vocabulary terms you created earlier ('France', 'Benelux', 'Switzerland') do not appear. They belong hierarchically to the department 'Purchasing'.

Department < x >

Region < x >

For the department 'Production' you could now create the values 'Switzerland,' 'Germany' and 'Italy,' for example, and select 'Italy' for correct keywording if you like.

➡ 1:N fields can extend over more than two hierarchical tiers. You could, for example, define a 1:n relationship between the field 'Region' and a new field 'Contact.' By doing this you would have created a third tier.

Field name	Contact
Field type	1:N
Link with field	Region
Label	Contact
Position	
Width	

Similarly, you could make several 1:n relationships for one field type 'Definition.' You must simply see to it that the field type 'Definition' is top of the hierarchy.

This is at the same time the most pragmatic approach: move top - down when you want to create hierarchically dependent fields. Always start with the field that is at the top of the hierarchy.

13.3.5 Number code

You can create fields which are linked to other fields. One alternative is the linking of a text field with a number code field.

13.3.6 Multi

The idea behind the field type 'Multi' is that with its help you can record several keywords of the same category for a document. Let us assume that you want to assign several contact persons per document. In the following we show you how to do it.

Go to 'Field definition' and create three fields with the names 'Contact', 'Contact1' and 'Contact2'. The field 'Contact' is assigned

the field type 'Definition', 'Contact1' and 'Contact2' receive the field type 'Multi' and the entry 'Contact' under 'Link with field'.

Field name	Contact1
Field type	Multi
Link with field	Contact
Label	Contact1
Position	0
Width	0

You can now allocate several contact persons per document. All values predefined under 'Contact' also appear in the combo boxes of 'Contact1' and 'Contact2'. See the example below:

Contact	Müller	< x >
Contact1	Huber	< x >
Contact2	Keller	< x >
	Müller	
	Huber	
	Keller	

13.4 Link with field

Here you enter the field to which the currently active field refers. For 1:N fields you choose the field with the superordinate concept (in Archivista terminology: 'Definition'), i.e. the field from which the others depend. In the case of multi fields, which are in actual fact copies of a main field, the main field is selected.

Field name	Region
Field type	1:N
Link with field	Department
Label	Region
Position	
Width	

13.5 Label

Do you want to attach a label to a field? If yes, please enter the label here.

➡ A label makes sense in most cases. With multi fields or hierarchically dependent fields it may be better to omit labels, though.

13.6 Position

Here you can specify at which position the field is to be displayed in the table. Use a number larger than '0'.

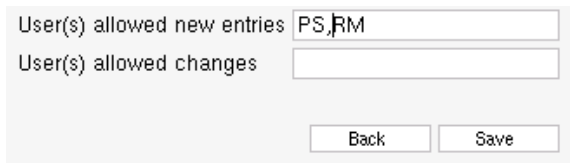
13.7 Width

Here you can specify the column width the field should take in the main view table. One screen pixel corresponds to 15 units, i.e. if you enter 900, the field will be 150 pixels wide.

➡ A value of '-1' means that the field in question is not to be displayed in the table at all.

13.8 User(s) allowed new entries

With the last two options in this dialog box you can lock a field for users not listed here.



The screenshot shows a dialog box with two text input fields. The first field is labeled 'User(s) allowed new entries' and contains the text 'PS, RM'. The second field is labeled 'User(s) allowed changes' and is empty. At the bottom of the dialog box, there are two buttons: 'Back' and 'Save'.

The users whose names are specified in **User(s) allowed new entries** can add **controlled vocabulary** to the pulldown list. This option concerns primarily 1:N or multi fields.

➡ Pay attention to separating the user names by commas.

13.9 User(s) allowed changes

Under **User(s) allowed changes** list those users who are allowed to make changes to this particular field. This does not concern changes to controlled vocabulary but solely changes to **selecting from** controlled vocabulary or changes to fields for which entries are not predefined.

➡ Pay attention to separating the user names by commas.

14 Archive administration

Options for archive administration can be set globally for Web-Client and database administration.

14.1 Fields in the archive

14.1.1 Hide extended icons

When you log in to the WebClient, two icons are displayed. One of these is for scanning and the other is for uploading files. If you do not want these icons to be displayed, you have to activate this option.

14.1.2 Number of columns for fields (1,2,3)

Here you can specify whether you want to display 1, 2 or 3 columns in detail view.

➡ The appropriate number of columns is only displayed if the browser window is sufficiently wide. Should this not be the case, the number of columns is reduced: from 3 to 2 or 1, from 2 to 1 column.

14.1.3 Start with reduced table

The table width can be reduced to the width of the view of the current file in the detail view. At the same time, the page (image)

on the right hand side of the detail view is displayed in full height (i.e. considerably enlarged).

14.1.4 Hide internal fields in reduced table

The fields 'Document', 'Date', 'Archived' and 'Folder' are hidden in the reduced table view.

14.1.5 Suppress field 'Title' in main view

In **Suppress field 'Title' in main view** you can globally define (i.e. for all users) if the field 'Title' is to be displayed or not. A modification takes effect only at next login.

14.1.6 Width of field 'Title' in table

Here you can determine the column width that the title takes in the table of the main view.

14.1.7 Field for publishing

The field that you choose under **Field for publishing** is a kind of second owner field with which a document can be allocated to a second owner (user or group user). For many archives this function is irrelevant, however, it can be a very important function indeed. Look at the following instance:

A knowledge archive is continually fed with new documents by several project managers. Each project manager is responsible for the documents she adds. At the same time she can make the documents available to an extended circle of users.

After this option has been activated, the input mask will show a second combo box with document owners. This way each primary owner of a document can set a second, subordinate owner. It

goes without saying that the main user settings apply also in the extended owner environment.

➡ The field that you specify here must be of the type 'Definition'. See 13.3.3.

14.1.8 Versioning

The following three functions 'Save file name in field', 'Versioning' and 'Field for key versioning' functions are explained in 14.4 and are only required if you require versioning, otherwise they remain empty.

Save filename in field	<input type="text"/>
Versioning	<input type="text"/>
Field for version key	<input type="text"/>

14.1.9 Elements in combobox fields (0-x)

The field `Elements in combobox fields (0-x)` gives you the possibility to restrict the number of documents displayed after searches in the WebClient. To define the number of search results displayed enter it in the **combobox field**. Thus, the **combobox field** simplifies and speeds up the search for a specific document in the **WebClient**. (Those interested in getting to know more about searching we refer to Chapter 8.2).

14.1.10 Show note field

If this option is active, a note field is displayed in the detail view.

14.1.11 OCR recognition

OCR-Erkennung	FineReader (ArchivistaBox AddOn)
PDF-Dateien erstellen	<input checked="" type="checkbox"/>
Gesamte Akte in eine PDF-Datei	<input checked="" type="checkbox"/>
Volltextabfragefeld beim Suchen einblenden	<input checked="" type="checkbox"/>
Seitentext in Ansicht darstellen	<input checked="" type="checkbox"/>

The **OCR** software serves to recognize text in the case of scanned text. Here you have a choice of four possibilities:

1. Leave the little box unticked if you do not wish any text recognition.
2. Choose **Commercial...** if your aim is to have the best possible results
3. Choose **Tesseract** if you wish to use this OpenSource technology. With this option you cannot create searchable PDF files. However, you have access to Gothic print recognition when you use language 'DeutschNeu'.
4. **Cuneiform** is another OpenSource text recognition software. It allows for the creation of searchable PDF files.

➡ Both OpenSource products know umlauts, the Tesseract engine can recognize only one language at a time, though (English, German, French, Italian, Spanish and Dutch). That is why with Cuneiform and Tesseract only the first language from the OCR definition (see 17) is taken into account. In the case of ArchivistaBox 'Dolder' only the commercial OCR option is available.

14.1.12 Create PDF files

Ticking **Create PDF files** means that the OCR run does not only make the ASCII text available for fulltext searches but in addition generates searchable PDF files. This option does currently not work for Tesseract 2.0.

14.1.13 Image compression in PDF files (1-100%)

Determine here at which quality the images are saved in the PDF files (1 = lowest quality, 100 = highest quality). A value of 10 or 15 should be sufficient to achieve very good results at reasonable memory requirements.

14.1.14 Whole document in one PDF file

By turning on the option '**Whole document in one PDF file**' you can effect that upon creation of PDF files each individual page is turned into a PDF file. Normally you should not turn this option off unless you must process extremely large documents with color images. In such a case it could be that the PDF file covering all pages would take more than 512 MByte space and could not be saved in the database because at the moment a maximum of 512 MByte (exception: ArchivistaDolder with 128 MByte) can be stored per database field.

14.1.15 Alternative barcode recognition

If you have purchased the module 'Barcode Recognition', there are two technologies at your disposal. Should any problems occur, switch to the alternative barcode recognition technology here.

14.1.16 View fulltext field (search form)

Using 'View fulltext field (search form)' you may specify whether to display the fulltext field in the mask in the WebClient and whether you wish to use MySQL or Sphinx for the full text search in your database. The latter can handle large volumes of data more quickly and is therefore highly recommended for large databases.

➡ **Important! Once Sphinx is enabled, this cannot be undone later.**

14.1.17 Show page text in 'View mode'

The function **Show page text in 'View mode'** enables you to see the text of the currently active document in the **WebClient's View mode**.

edit	Land
	Region
	FirmenNummer
	FirmenName
	Dokumententyp
Metabo (Schweiz) AG Ringstrasse 30 CH-8317 Tageiswangen	
Tel 052 3543434 Fax 052 3543435	
Service de réparation et pce de rech. Tel 052 3543444 Fax 052 ,3543445	
Metabo	
COPIE	
Diserens Electromécanique Route de Marcolet 45 1023 Crissier	
Diserens Electromécanique Route de Marcolet 45 1023 Crissier	
Bulletin de livraison ^einfo	

This function is exceedingly useful if one wants to copy the entire text of the page without any graphics, barcodes or similar.

One should turn off this function if one wants to put the documents on the Internet: Flaws in text recognition can be distracting.

14.2 Options for page text and search

Copy fields to page text	<input type="checkbox"/>
Hide field 'Owner'	<input checked="" type="checkbox"/>
Startup with photo mode	<input type="checkbox"/>
Show download link	<input checked="" type="checkbox"/>
Keep search values	<input type="checkbox"/>

14.2.1 Copy fields to page text

In full text retrieval only the recognised page text is searched. If you turn on this option, all indexed fields are copied to the page text (they are added to the first page of a document). Thus, a full text search returns also hits that stem from keyword attribution and you even do not need to know in which field a specific keyword was entered.

14.2.2 Do not show owner

Hide field 'Owner'	<input type="checkbox"/>
Startup with photo mode	<input type="checkbox"/>
Show download link	<input checked="" type="checkbox"/>

Activation of the field **Do not show owner** has the effect that the field **Owner** does no longer appear in the **View** mode of the **WebClient**. The field serves to indicate which user or user group the currently active document belongs to. To which documents a user has access is described in detail in Chapter 11.1.2. Often the document owner is of no relevance. In this case the field need not be displayed.

14.2.3 Startup with photo mode

If you tick **Startup with photo mode**, then the photo mode is active when you login to the database, i.e. in the upper part of the screen you will see the respective first pages of the documents instead of the table. This view makes most sense for photo collections.

14.2.4 Show download link

When you access the **WebClient** and the field **Show download link** is active, the table shows the field **Download** at the end of each row.

This field offers the possibility to download the currently active document as PDF file. To be able to download a PDF file the function **Create PDF files** need not necessarily have been activated. One can create a PDF file ad hoc. However, if you want PDF creation to happen automatically and constantly in the background and if you want searchable PDFs you must not only tick the box **Create PDF files** but also **Show download link**.

14.2.5 Keep search values

By activating this option the search parameters of the last search appear each time you go to the search form.

14.3 Options for archive

Quality factor JPEG (1-100%)	33
Compress JPEG files always	<input type="checkbox"/>
Scaling of preview (0,10-100%)	0
Ignore duplicate images on camera import	<input checked="" type="checkbox"/>
Save image files in source format	<input checked="" type="checkbox"/>
Conditions for archiving process (SQL)	
Folder size external storage (MByte)	30
Max. count files in folder	50
Size of CD/DVD? (MByte)	600
Log database access	<input type="checkbox"/>
Check hash to deactivate	

14.3.1 Quality factor JPEG (1-100%)

Quality factor JPEG (1-100%)	33
Scaling of preview (0,10-100%)	20

Color scans are saved in JPEG format. With this option you can determine the quality of your images. 100% is best quality. The smaller the value you enter the higher the compression.

14.3.2 Compress JPEG files always

With this option turned on, JPEG files are always compressed to the quality specified. Normally, this option should not be active, unless incoming JPEG files (e.g. from a network scanner) are so large as to overly stress the archive.

14.3.3 Scaling of preview (0, 10-100%)

Here you can enter the relation of the size of the preview image to the original file. 25 for 25% is a good value to start with. '0' means that no preview images are created at all. In general, however, it is good to activate 'Preview'; access to the images is much faster even if this option at 25% requires approximately 10 % more disk space.

14.3.4 Ignore duplicate images on camera import

When you connect a camera or a stick to a USB port of ArchivistaBox, an image import dialog will pop up (cf. 32.7). Activate this option in order that you may not import the same images into ArchivistaBox several times. On each import a key (MD5 string) for clear identification is created. This key is saved in the Archivista database.

During any later import, every image is checked for duplicates on the basis of the key. If there is already a copy of the image in the database, then the image is not imported again. In other words: Images are imported only if they do not already exist in the database. When this option is turned on, the database creates an MD5 key for all pages. This will take some time (ca. 1000 pages per minute). As soon as all pages are provided with a key, a message will be saved in the log file (see section 24.4.2 in WebConfig).

14.3.5 Save image files in source format

Normally (= if this option is not activated), all colour images are saved in jpeg format. Image files cannot be checked out (in contrast to office and PDF files). The original image files are not saved in the archive because they would need a lot of additional space.

If you activate this option, then the image files are saved in the original file format, too. You can check out and again check in these files, even versioning is possible if versioning is activated.

14.3.6 Conditions for archiving process (SQL)

Conditions for archiving process (SQL)	<input type="text"/>
Folder size external storage (MByte)	<input type="text" value="300"/>
Max. count files in folder	<input type="text" value="1000"/>

By means of an SQL fragment you can specify that the archiving process should only involve certain documents and not others. As example may serve `Publish_<>_'`. This means that only the documents published to other users should be part of the archiving process.

14.3.7 Folder size external storage (MByte)

At regular intervals an Archivista archive undergoes an archiving process. This automated process 'rearranges' and compresses the newly added pages and generally does everything that needs to be done. One important point in this is the preparation of pages for the copying to external media. I.e. in order that you may save your data to a CD or DVD the system prepares a folder of the appropriate size. Under **Size of folder for external storage (MByte)** you specify the maximum size. If you work with CDs, for example, you might enter 600 here.

14.3.8 Max. count files in folder

In addition to indicating the maximum folder size also the maximum number of pages that are supposed to fit in a directory (folder) after the archiving process can be determined.

14.3.9 Size CD/DVD (MByte)

This option lets you indicate what size the external medium to which you will copy your data will have.

14.3.10 Log database access

Log database access	<input checked="" type="checkbox"/>
Check hash to deactivate	<input type="text"/>

This option enables you to have a log created for all access activities regarding the archive. You find more information on this in 25.7.1.

➡ This option results in the ArchivistaBox working a bit more slowly. Accordingly, turn it on only if you attach importance to having an access log. What is more, you should be aware that the option cannot be turned off so easily once it is turned on, the reason being that otherwise an administrator could hide his or her own accessing the database.

14.3.10.1 Turning off the access log function

Should the access log function accidentally have been turned on or should you wish to deactivate it for other reasons, you can do it by unticking the option 'Log database access' and entering at the same time an 8-digit checking number or check hash. You can take the checking number from the 'access log' table. It corresponds to the first eight characters of the first generated hash (field 'hash').

14.3.10.2 Check integrity of access log

An access log is of little value if we cannot make sure that the log itself cannot be falsified. For this reason keys are generated for each entry in table 'access', keys that later on may serve to confirm the accuracy of the entries. The technology used here was taken from the application 'md5sum'. When so-called 'signing' of the entries takes place, it is always the entries that are signed as well as the hash of the previous entry. This is the reason why it is possible to ascertain later on whether all entries were correctly created. But only if the hash of the first and last (most current) entry are known. Consequently, we recommend that you save the first hash in a different place after you have turned on the access log function, and to save the most current hash from time to time too.

➡ If you have a good backup concept, it is unnecessary to manually note the hash as these keys are automatically saved in the backups.

14.4 Enabling versioning

As mentioned, versioning must be enabled first. To do this, log into the WebAdmin. For versioning you need three fields: one field (Versioning) to administer the document root, a further field for the version number (VersionNo), and it is recommended to reserve a third for the file name (FileName).

Field definition @ archivista

>New	Field name
>Edit >Delete	VersioningKey
>Edit >Delete	Versioning
>Edit >Delete	Filename

The fields thus created must be assigned to the masks on which the versioning is to occur. Add the three fields, 'Versioning', 'VersionNo' and 'FileName', to the 'Masks' menu item.

Mask definition @ archivista

Mask: English
English

>New	Field name
>Edit >Delete	VersioningKey
>Edit >Delete	Versioning
>Edit >Delete	Filename

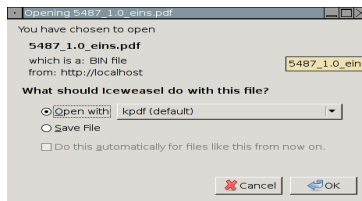
Then you need to enable and manage these fields in the archive.

Archive administration @ archivista

Suppress field 'Title' in main view ☐
Hide extended icons ☐
Width of field 'Title' in table 3000
Field for publishing
Save filename in field
Versioning
Field for version key
Elements in combobox fields (0-x) 0
Show note field ☐

Versioning is thereby enabled. You can then upload Office files either via the WebClient or an FTP upload.

➡ **Important! No checking in or out of the ArchivistaBox is necessary for versioning. Whether this is on or off can be established as follows.**



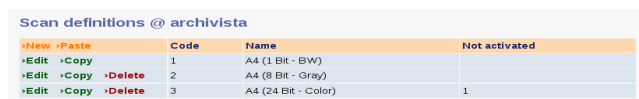
Once you have accessed a file with active versioning from the WebClient, you will see the file not only with the file name, but preceded by a file stamp. In the above example, the name is 'eins.pdf'. When you first check out, you will receive the file name `Doc_Version_Name.pdf`.

To learn how to use versioning, consult section 9.13.

15 Scan definitions

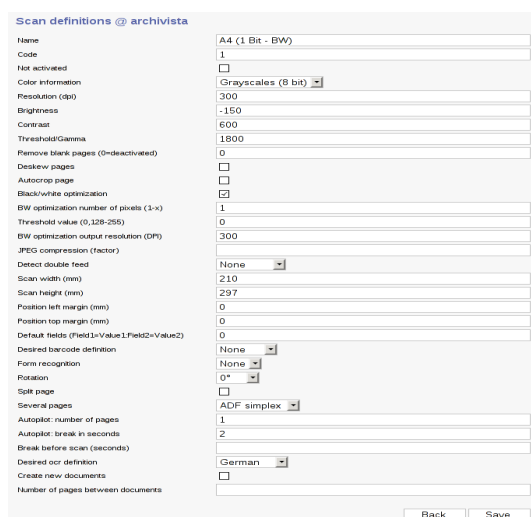
15.1 Introduction

Here you can create those scan definitions which you are likely to need again and again. After clicking the 'scan' menu item you will first see a list of available scan definitions. As of the 2010/III release, scan definitions can also be disabled temporarily and provided with scan codes.



	Code	Name	Not activated
New Paste			
Edit Copy	1	A4 (1 Bit - BW)	1
Edit Copy Delete	2	A4 (8 Bit - Gray)	1
Edit Copy Delete	3	A4 (24 Bit - Color)	0

Changes in scan definitions or their deletion is done by means of the 'Edit' and 'Delete' functions marked in green and red, respectively. In the first upper left-hand box of the table the word 'New' appears in red letters. Click on 'New' to create a new scan definition. You will see the following screen:



Scan definitions @ archivista

Name: A4 (1 Bit - BW)

Code: 1

Not activated: ☒

Color information: Grayscale (8 bit)

Resolution (dpi): 300

Brightness: -150

Contrast: 600

Threshold/Gamma: 1800

Remove blank pages (0=deactivated): 0

Deskew pages: ☐

Autocrop pages: ☐

Black/white optimization: ☒

BW optimization number of pixels (1-4): 1

Threshold value (0, 128-255): 0

BW optimization output resolution (DPI): 300

JPEG compression (factor):

Detect double feed: None

Scan width (mm): 210

Scan height (mm): 297

Position left margin (mm): 0

Position top margin (mm): 0

Default fields (Field1=Value1,Field2=Value2): 0

Desired bar code definition: None

Form recognition: None

Rotation: 0°

Split page: ☐

Several pages: ADF simplex

Autoplot: number of pages: 1

Autoplot: break in seconds: 2

Break before scan (seconds):

Desired ocr definition: German

Create new documents: ☐

Number of pages between documents:

Back Save

15.2 Settings

15.2.1 Name

Enter the name of the scan definition here, e.g. 'A5 color landscape'.

15.2.2 Code

You can define a code for every scan definition. You can use this code later to scan with the according key (i.e. 1+Enter).

15.2.3 Not activated

You can deactivate a scan definition temporarily. In the overview of scan definitions there will appear '1' in column 'Not activated'. This way, you do not need to delete rarely used scan definitions and retype them later when they become relevant again.

15.2.4 Color information

Here you specify the colour depth, i.e. you decide if the page is to be scanned in black and white, grayscales or color.

15.2.5 Resolution (dpi)

Under **Resolution** you indicate the scanning resolution. It goes without saying that your scanner must support it. Normally you use 300dpi.

15.2.6 Brightness

Normally, the value here is 0. If you have to scan very bright or very dark pages you can increase or decrease brightness.

15.2.7 Contrast

Also for contrast the standard value is 0. If you want to increase contrast, try entering 10, 30 or 50. If you want to decrease it, use a minus value.

15.2.8 Remove blank pages (0=deactivated)

This function makes most sense whenever you want to scan your papers recto and verso although only a part of the stack (one third, for example) has any information on the reverse side. Enter the conditions for the removal of blank pages in the following way:

20.56

The first is a threshold value for the amount of black on the page expressed in one-tenth of a percent. '20', for example, means that any page whose surface area is less than twenty-tenths of a percent (or 0.02 percent) black will be deleted. (The margin mentioned above is not counted in the calculation of this.)

The second value stands for the margin in pixels that is **not** taken into consideration for the calculation of the proportion of black (as against white) on the page.

➡ Please note that the two values need to be separated by a full stop. We recommend that you take a figure between 10 and 40 for the first value. The higher the first value the more black a page may contain. The lower the first value, the higher the chance that pages with only a little black information are integrated into the archive. The second value must be '8' or a multiple of it such as '40', '48', etc.

0 means that there are no empty pages to delete.

15.2.9 Black/white optimization

When this function is activated, scanning takes place in greyscales. As soon as the images arrive at the ArchivistaBox they are turned

into black and white by means of an optimization process. I.e. contrasts are reinforced so that the text remains legible even if the original page was of low print quality.

15.2.10 Threshold (0, 128-255)

With this value you determine how the images are converted back to a black/white raster after the optimization process. The higher the value the higher the proportion of black. If you enter '0', the software will try to remove a colour background.

15.2.11 BW optimization number of pixels (1-x)

The value entered here has an impact on the black/white optimization. It determines the number of points the algorithm reads around the point of origin. The best way to proceed is to test a few typical documents. '3' is a good value to start out from.

15.2.12 BW optimization output resolution (DPI)

Here you determine the resolution the image must have when it is stored in the database. A typical value would be '300'; it is generally recommended for black and white pages.

15.2.13 JPEG compression (factor)

Between scanner and computer (ArchivistaBox) image data are normally transmitted without having been compressed. At 300dpi and colour there are ca. 25 MBytes of data per page that must be conveyed. With double sided scanning there are about 50 MBytes per sheet. The USB interfaces used today can in theory handle 60 MBytes (or 480 MBits) per second. In practice the throughput is only about half of that, though, and data transmission may easily take a couple of seconds. There exist a few scanners today which compress the image data while they are still in the scanner

and which send out an already compressed data stream. The transferral of data between scanner and computer is much more efficient this way. If the device you use supports such compression, you can enter the compression factor here.

➡ Please note that currently only a few Fujitsu scanners offer this option. Use 0 (no compression) or choose a value between 1 (fast transferral with less quality) and 7 (slow transferral, optimum image quality).

15.2.14 Doublefeed detection

Some Fujitsu scanners offer ultrasonic doublefeed detection. By clicking this check field you can activate it. The ultrasound helps the scanner to notice if there is really only one sheet being pulled through. This helps to avoid paper jams and erroneous feeding of more than one page.

➡ Please note that to date only a few Fujitsu scanners support this option.

15.2.15 Scan width (mm)

The scan width of an A4 page is 210mm.

15.2.16 Scan height (mm)

The scan height of an A4 page is 297mm.

15.2.17 Position left margin (mm)

This value is needed with some scanner models when the width of the page does not tally with the area the scanner covers. For example when you use the ADF to scan pages that are narrower than US letter, it is not sufficient to change the scan width but the left margin must be adjusted too.

15.2.18 Position upper margin (mm)

With some scanner models it may be necessary to adjust the upper margin when scanning special paper sizes. This can be done by means of this function.

15.3 Post editing

Postediting options for scanned pages are available. In the following these options are explained.

15.3.1 Default fields

You can enter default values for specific fields here. These fields are then automatically filled during scanning. The following format must be adhered to:

`Field1=Value1:Field2=Value2:script:Field3=Value3`

15.3.2 Calling a program during scanning

You normally always use `Field name=Value` followed by a colon before the next entry. In addition to fixed fields, you also have the `script` option. Here, the relevant program is called after each page that is scanned.

This program is transferred via the command line to the file number, host (computer), database, user and password of the active file. The program can then register in the system and carry out specific post processing activities. For example:

```
#!/usr/bin/perl

my ($host,$db,$user,$pwd,$lnr) = @_;
use DBI;
my $dns = "DBI:mysql:host=$host;database=$db;";
my $dbh = DBI->connect ($dns,$user,$pwd);
```

```

if ($dbh) {
    my $sql = "select client from archive where file number=$lno";
    my @row = $dbh->selectrow_array($sql);
    my $sql1 = "";
    if ($row[0] == 300) {
        $sql1 .= "Owner='HHWRW'";
    } elsif ($row[0] == 100) {
        $sql1 .= "Owner='KISRW'";
    }
    if ($sql1 ne "") {
        $sql1 = "update $db.archive set $sql1 where file number=$lno"
        $dbh->do($sql1);
    }
}
}

```

The above script adjusts automatically to the 'Owner' field if the 'Client' field adopts a value of 100 or 300.

The corresponding script was entered here, for example, with `auto.pl` in the input fields. It must be stored under the following path in the ArchivistaBox:

```
/home/data/archivista/cust/autofields/auto.pl
```

➡: Logically, the scripts are linked with the barcode recognition in the input fields. You even can administrate the jobs in WebAdmin, see 22.

15.3.3 Desired barcode definition

When you enter a barcode definition created earlier 16.4, the barcode recognition in question will be carried out automatically after scanning. This way it is ensured that you can work with more than one barcode definition and the document is 'steered down the right path' from the moment the scan button has been touched.

15.3.4 Rotation

Here you can specify that the page undergoes automatic rotation after scanning. This makes sense for pages with a landscape orientation, for example. Choose from the values 0, 90, 180 and 270 degrees.

15.3.5 Split page in middle

This function serves its purpose when books are being scanned. For more information see Chapter 32.4.

15.3.6 Several pages

If you want to scan more than one page at one go there are a few options to simplify your task.

- ADF Simplex: simplex scanning with ADF
- ADF Duplex: duplex scanning with ADF
- ADF 270/90: duplex scanning with ADF where recto has to be rotated 270 degrees after scanning and verso 90 degrees in order that all pages 'stand on their feet' at the end of the process
- Autopilot: manual scanning with the help of the flatbed scanner; here there are two further settings

15.3.7 Autopilot: number of pages

If you plan to scan a book, for example, you can make the (flatbed) scanner scan at regular intervals while you have to turn the pages only. Enter here the number of times you want the scanner to repeat the scanning process.

This function is currently not implemented. The flatbed mode can be selected but only one page at a time is scanned.

15.3.8 Autopilot: break in seconds

Here you indicate how long the scanner should wait before doing the next scan.

15.3.9 Break before scan (seconds)

If the scanner is located not on your desk but further away you can enter delay time here.

15.3.10 Desired OCR definition

Choose here the OCR definition which should be applied to the pages scanned with this scan definition.

15.3.11 Create new documents / Number of pages between documents

These two points belong together. Here you specify that a new document must be created at regular intervals and after how many pages this must be done.

➡ Example: Let us assume that you scan invoices most of which consist of one page. By activating 'Create new documents' and entering '1' after 'Number of pages between documents' you can build a scan definition that allows you to put a whole stack of paper into the scanner: after each page a new document is created automatically. The scanned 'stream' of pages is thus separated into individual documents without the help of barcode technology. If you want to scan invoices that come as two pages, you can create a second scan definition (and enter '2' beside 'Number of pages between documents').

16 Barcodes (Archivista Box)

16.1 General

16.1.1 Barcode processing with Archivista

The purpose of the Archivista barcode solution is automated handling of (scanned) documents that are marked or labelled with barcodes. There are three main areas to consider: barcode entry and labelling, barcode recognition, and analysis of the recognized barcodes. After a short introductory part these three areas form the core of this chapter.

➡ A prerequisite for the use of this function is that the barcode module has been purchased and activated.

16.2 Barcode technology

16.2.1 Barcode processing: why?

Scanning is only the first step when keeping an archive. Fast retrieval of your data when you need them is equally important. Archivista 2014/IX offers you both full text retrieval and the entry of keywords.

Full text retrieval is especially useful for archives which lend themselves well to random searches by individual words or word combinations. Working with keywords makes sense in all cases where the filing system in the computer adheres rather closely to a classical paper filing system. Barcode processing goes a step further by efficiently handling large masses of stereotyped documents.

To put it simply, we mark out each document with some unequivocal piece of information according to which a later search may be executed. We achieve this by adding a reference number to our document before we print it or by sticking a barcode label (with

reference number) on the document to be filed. As a consequence each document carries a barcode and reference number when it is scanned.

Scanned pages are subsequently worked on by a barcode recognition software and from now on they are at our disposal for searches by reference number or further data analyses. An enormous time saving potential is realized if documents are filed and searched with the help of barcode technology.

16.2.2 What is a barcode?



Barcodes consist of black bars and white spaces. The sequence of these bars and spaces differing in width serves to encode information which can be read (by a scanner) and processed (by a computer) in an automated fashion. The relatively simple encoding allows barcode reading devices to recognize the information with 100% accuracy.

16.2.3 Which barcode types must be distinguished?

There is quite a variety of formats in use. Common to all of them is their (for lay persons undistinguishable) appearance of black bars and white spaces of different widths. We differentiate between two types of barcodes: two level codes and multilevel codes.



Two level barcodes have two different widths: narrow and wide. Interleaved 2 of 5, MSI Code, Codabar, Code 39 and Extended Code 39 are examples of two level codes.



Code 128

Multilevel barcodes have more than two variants of width, they may have bars and spaces that are very narrow, narrow, normal, wide and very wide, for example. Among the multilevel barcodes we count Code 11, UPC/EAN and kin, Code 128 and Code 93.

16.2.4 Which barcodes are supported by Archivista?

A barcode solution for the purposes of document management must be able to allocate barcodes to documents and vice versa. To follow this purpose Archivista offers the most widely known barcodes but does not support each and every barcode format. The Archivista Barcode Option works with the barcode formats Code25, Code39, Code128, EAN8 and EAN13. The auxiliary program BarcodePrint supports Code25, Code39 and Code 128. Out of this reason these three formats are presented below:

Code 25: numeric only barcode, very efficient. Normally you use this format because it can represent also higher numbers in limited space.



Code 39: alphanumeric barcode. Its ability to represent not only figures but text as well makes it more space consuming than the previous solution.



Code 128: alphanumeric barcode. A space saving multilevel format which yields relatively narrow, unobtrusive barcodes even if figures and text are encoded.



16.3 Barcode entry

16.3.1 Barcodes exist on documents to be scanned

If the barcodes are affixed to the documents during the printing process, you can scan and analyze the documents with the help of Archivista 2014/IX and its barcode application without further ado.

16.3.2 Working with Archivista-BarcodePrint

If the documents to be filed do not yet carry a barcode, the barcode information must be added (normally by sticking on a label). This must take place before scanning, otherwise no barcode processing is possible!

To do this, we recommend the use of the auxiliary program BarcodePrint, which we will be pleased to put at your disposal and of which we show the main mask below:

16.3.2.1 Barcode entry

Working with Archivista-BarcodePrint is simple. After starting the program you get to the entry mask. Now enter the barcode information in the **'Entry field'** and the corresponding barcode becomes visible below. By pressing the Return key the entered reference number is copied to the list for printout on the right. The field 'text for barcode' is ready for the next reference number which again you copy to the list on the right by hitting the Return key. When all reference numbers are entered you can print the barcodes by clicking the field 'Print'.

➡ Barcodes that have been entered previously may be called up once again by clicking the combo entry field (black triangle on the right).

16.3.2.2 Editing and deleting individual barcodes

Entered barcodes are moved to the 'list for printout'. There you can edit and delete them if you wish. To do this, click the area **'List for printout'** and correct or delete the entry in question.

16.3.2.3 Copying a barcode to the clipboard

It is possible to copy the barcode to another Windows application. Simply click the field 'Copy' after entering the barcode information and paste it to any Windows application which is able to handle bitmap files.

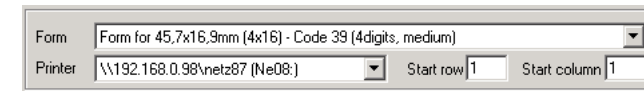
16.3.2.4 Printing of entered barcodes

To print all barcodes appearing in the list click 'Print'. All barcode information generated and visible in field 'Entered barcodes' will be transferred to the printer.

➡ Printing takes place without further confirmation!

16.3.2.5 Printing options

If Archivista-BarcodePrint has once been adjusted to your needs, you normally do not have to change any settings. However, once in a while you might need the printing options to achieve the desired printing result.



First, choose the appropriate form. The form active during the last session normally appears as standard form at the beginning of the next session. You create new forms by clicking the button 'Forms'. You can also edit forms, for example, to adjust the page layout.

Choose the appropriate printer under **'Printer'**. After starting up Archivista-BarcodePrint it is always the standard printer that appears.

The options **'Start row'** and **'Start column'** are needed to help you saving label sheets. Printing starts only in the mentioned row and column.

16.3.2.6 Using label sheets

To achieve best results barcodes should be printed with a laser printer, which gives the bars a sharp edge. If the barcodes are printed in a large format an ink jet printer may also serve the purpose.

Ordinarily, you use self-adhesive label sheets. By default Archivista-BarcodePrint works with the label sheets from the Zweckform company: **Zweckform No 3667, 4732 and 6071**. The standard forms were created for these label sheets.

16.3.2.7 Deleting all added barcodes

After printing the barcodes the individual barcode entries are still in the printing list. You can print them again or edit them if you like.

If you do not need them anymore you delete them by clicking the field 'Delete all'. All entries in the printing list are erased.

16.3.2.8 Importing existing barcode information

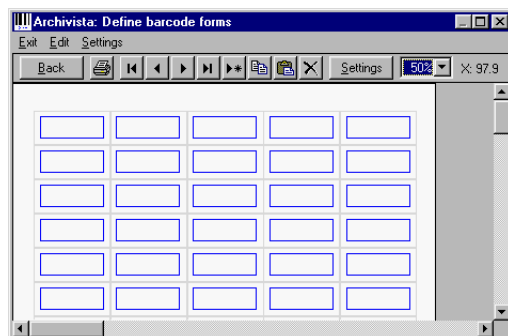
If barcode information exists already in electronic form you can import it via an ANSI file directly into Archivista-BarcodePrint. Use the button 'Import'.

16.3.3 Archivista barcode forms

If you cannot print your barcodes by means of the existing forms, you can edit these forms or create entirely new ones.

16.3.3.1 Form generator

From the main mask you get to the form generator by clicking the button 'Forms'. You will see the following:



All functions with regard to adding and deleting individual forms you find in menu 'Edit'.

➡ The chosen form is shown on screen. Each little blue box represents a barcode field.

16.3.3.2 Making form definitions

You can define what a form looks like by going to the **form generator** in menu 'Settings' / 'Page settings' or directly by clicking the button '**Settings**' in the icon bar. Now you can enter the settings according to your needs.

Under '**Form name**' you give a name to your form. This information appears later in the main mask where you can choose among the different forms defined by you.

Under '**Page setup**' you find all the necessary page information. Please note that the **page size values are not sent to the printer** when printing. The advantage is that you can print with all paper sizes. You must ensure that the values set do not exceed the physical paper size, though. With some printers it may be necessary to move the whole page a little to the left or to the right, or a little up or down. To do this you use '**Adjust left margin**' and '**Adjust upper margin**'

Under '**Main area**' you give the number of rows and columns as well as the margins for the individual barcode cell.

Under '**Barcode type**' you choose the type of the barcode to be printed. Choose between the numeric only barcode 'Interleaved 2 of 5' and the alphanumeric barcodes 'Code 3 of 9' (= 'Code 39') and 'Code128'.

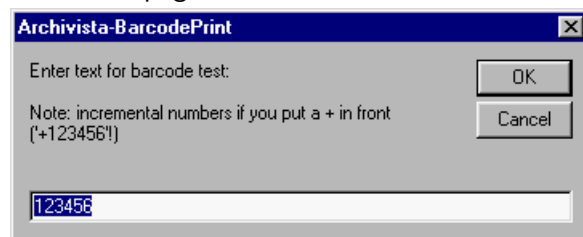
Under '**Size of barcode**' you define how the barcode looks. You can choose between three standards: '**small**', '**medium**' and '**large**'. If you want to define your own choose '**special**'. The fields '**barcode height**' and '**spacing**' are at your disposal. 'Spacing' influences the gaps between the black bars and the white spaces. You can see the changes you make in the barcode picture on the right of the entry fields. **If you do not want to print the barcode text**, tick the corresponding box.

16.3.3.3 Saving of form settings

Changes to forms are registered in the memory after you confirm your changes before you go to another form or when you leave the form generator. To save the changes and have them ready for further sessions you must end the program. Only at that point the new form formats are written to the file 'AVBCODE.DAT'.

16.3.3.4 Print test page

After you have entered all parameters in the form 'Settings' you should print a test page in the **form generator** to check the looks of it. To do that either click the printer icon or choose the option 'Print test page' in menu 'Exit'.



Enter the text of one barcode as example and click 'OK'. After that a whole page with the same barcode is printed to give you the possibility to check margins, readability by scanning it, etc.

➡ To achieve incremental numbering you put a plus in front of the barcode information (e.g. '+123456').

Problem solving: if the printout does not give the desired result, you must alter the form settings accordingly. Problems occur often because the margins are set below the minimal printer margins. In such a case you simply increase the margin.

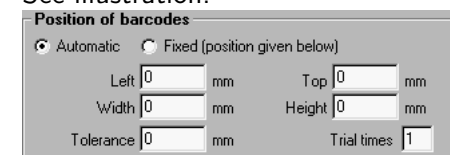
16.3.4 Sticking barcode labels on documents

After printing the label sheets with Archivista-BarcodePrint you stick the labels onto the documents. There are two ways to do that:

16.3.4.1 Labels in variable positions

Normally you work with this option. In principle you can affix the label to the document anywhere you like. However, you should take care to place it in a way that there are a few millimeters of space between label and surrounding text.

➡ If you work with this option you must tick it in 'Options for barcode recognition' ('Database', 'Parameters' and 'OCR settings'). See illustration:



Problem solving when barcode is not recognized: with the 'automatic' recognition of the barcode, i.e. when the recognition software locates the barcode itself, problems may occur when the barcode is not long enough. If you work with format 'Interleaved 2

of 5' and 6 digit barcodes this may be the case. You best change to the longer format 'Code 3 of 9' / 'Code 39' to avoid this problem.

16.3.4.2 Barcode labels in the same position

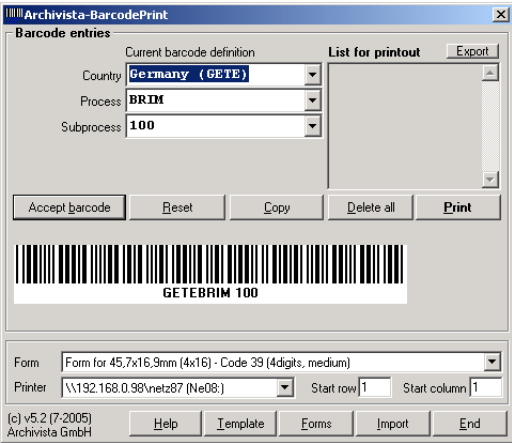
It may make sense to position the barcode label on all documents in the same place. In particular if there is other barcode information on the document or if the barcode is so complex that 'automatic' recognition presents problems.

➡ If you work with this option, you must tick it in 'Options for barcode recognition' ('Database', 'Parameters', 'OCR settings'). Choose 'fixed' and enter the details of the preferred barcode position.

16.3.5 Define your own BarcodePrint forms

It is not difficult to tailor the BarcodePrint form to your needs. However, we recommend that you study this subchapter at some calmer moment since the possibilities are rather comprehensive. Investing some time here may save you enormous amounts of time later on.

When you start the Barcode program after having installed Archivista 2014/IX you will find exactly one field into which a barcode may be entered. The illustration below shows you an example of how the BarcodePrint form may be adjusted:



You see three fields which together generate one barcode. The parameter file 'avbcode2.dat' serves to check whether the barcode is correct. This parameter file lies in the program directory of Archivista BarcodePrint.

16.3.5.1 Working with templates

Thanks to the button 'Templates' you can easily adjust the barcode entry form.

BarcodePrint							
	Label	Start	Length	Min.	Max.	Default	Value1;Value2;... OR Value1=Def1;Value2=Def2;...
Part 1	Country	1	4	0	9999		GETE=Germany;ITTE=Italy;CHTE=Switzerland
Part 2	Process	5	4	0	9999		BRIM,DSAM,HTAX
Part 3	Subprocess	9	4	0	9999		100;200;300;400
Part 4							
Part 5							

Start,End=Positions in barcode; Min,Max=Value range (-1=no range); Default is optional

Cancel Save

Each 'analysis unit' of the barcode takes one row. Take care to separate the values in the last field by means of semicolons.

Alternatively, you can adjust the entry form directly in the parameter file 'avbcode2.dat'.

Within a parameter file a maximum of five entries are possible because there are only five fields. The lines 2 to 6 of the parameter file are decisive. Each row is built along the same lines:

- **Label:** text that is supposed to appear in BarcodePrint form, e.g. Country
- **Starting position:** specifies what position this entry takes in the barcode; the first entry starts at '1'
- **Length:** the maximum length of the entry; a '3' means that this part of the barcode consists of three characters
- **Min:** minimum value of barcode
- **Max:** maximum value of barcode
- **Default:** the standard value (must be between 'Min' and 'Max')
- **Value1;Value2;... (optional):** here you can specify up to 100 precise entries that should be accepted. The format of this list is the following: 'Value=Description;'. For 'value' you enter the figures or letters that are part of the barcode, for 'description' the name you want that analysis unit to stand for.

➡ With the help of the parameter file 'avbcode2.dat' you may create barcodes for the formats 'Code25', 'Code39' and 'Code128'. Alphanumeric characters (i.e. with letters) are supported by 'Code39' and 'Code128'! Before you can create a new template you have to make sure that in the barcode **form** (press button 'Form') the right barcode type is specified.

16.3.5.2 Series

In a limited way it is possible to create barcode series.

Let us consider the following example. Twice a day goods are dispatched from your warehouse. The delivery notes are filed in a folder. At the end of the week you would like to scan the

documents but without detailed indexing information. Your aim is above all fast indexation. Only the date and a '1' or a '2' for the morning or the evening dispatch, respectively, are required. BarcodePrint enables you to create the necessary barcodes with one click. By entering the following information in the barcode print program a series of barcodes is automatically prepared:

The date can be entered in English format 'mm/dd/yy'. However, the date as it appears in the barcode displays the format 'yyymmdd'.

With one click barcodes for the whole week have been created: they carry the date followed by '1' or '2'.

The template belonging to the example above determines how the entry mask looks:

The code entered under 'Length', namely '8e', defines that the date will be printed out in 8 digits and that it can be entered according to Anglosaxon custom, e.g. 12/31/06.

The codes used are

- 6d
- 6e
- 8d
- 8e

whereby 'd' stands for German entry, 'e' for English entry and '6' or '8' for the number of digits needed for output.

16.4 Barcode recognition

Here you specify different barcode recognition definitions. After clicking the menu item 'Barcode recognition' you first see a list of existing definitions.

Barcode recognition @ testdb

>New	Name
>Edit	Barcode
>Edit	>Delete
	Barcode2

To make changes to existing definitions click 'Edit'. To delete definitions click 'Delete'. The first cell in the upper left hand corner shows 'New' printed in red. Click 'New' to create a new barcode recognition definition. You will see the following dialog box:

Barcode recognition @ testdb

Name	Barcode
1st barcode type	Code128
2nd barcode type	Code25
Alignment	Left to right
Check character(s)	
<input type="button" value="Back"/> <input type="button" value="Save"/>	

The system is basically designed for parallel use of different barcode definitions. The individual settings are described below.

The different definitions are distinguished by '**names**' that you enter here. Under '**1st and 2nd barcode type**' you specify the kind of barcode to be recognized. Archivista recognizes the following barcode types: Code39 (with or without test characters), Code25 (with or without test characters), EAN13, Code128 and EAN8. If type is set to 'automatic', every barcode type is recognized.

➡ The barcode recognition can simultaneously recognize two different barcode types, e.g. Code128 and Code25.

Further accuracy in barcode recognition can be added by setting the **barcode's alignment**. The options are 'left to right', 'bottom to top', 'right to left', 'top to bottom'. When set to 'automatic', any alignment will be recognized.

Check characters give even further control over barcode recognition. However, this function only works if all barcodes to be recognized begin with the same set of ID characters. By entering 'STAT,PERS', for example, you effect that only barcodes beginning with STAT or PERS are recognised as valid. This increases accuracy of barcode processing.

You can use **Stretch vertically** to stretch barcodes of less than one inch in height so that they reach the minimum height.

You use **Recognition after each page** to ensure that barcode recognition occurs even if the same barcode exists on a subsequent page. This guarantees that each page with a barcode is always stored in a separate folder.

At **Barcode engine** you can set which of the two barcode technologies you want to use for this particular barcode definition. 'Standard' uses the engine specified in 'Archive administration'. 'First engine' means that always the preset technology is used (alternative barcode recognition is not active). 'Second engine' always uses the alternative barcode recognition. Simply put: Should

the barcode recognition not work after testing with all parameters, one can conveniently define/test the alternative technology without it having a negative effect on the other barcode definitions.

16.5 Barcode processing

Here you specify how the barcodes are to be decoded after text recognition i.e. which part of which barcode is to be allocated to which field.

After clicking the menu item 'Barcode processing' you first see a list. Each line represents an 'analysis unit' or piece of information that is allocated to a specific field. Thus, the information of one barcode may be 'food' for several fields or, indeed, two barcodes may deliver information that fills as many as ten fields.

Barcode processing @ testdb

Barcode definition

	Barcode	Length	Field	Start	Characters
Edit Delete	1	12	Title	1	4
Edit Delete	1	12	Test	5	4
Edit Delete	1	12	Region	9	4

To delete definitions click 'Delete'. To add definitions or make changes to them click 'Edit'. You will see the following dialog box:

Barcode processing @ archivista

Barcode definition

Barcode

Length

Field

Start

Character

The individual settings are described below.

➡ Please note that what you enter here will later run in the background when you or your colleagues scan documents with barcodes on them.

Barcode definition: Choose here to which barcode the following settings apply. (Click the **Select** button.)

Barcode: Here you narrow down further to which of the five barcodes that theoretically can be gathered under the umbrella of a barcode definition the following settings apply.

Length: Enter here the total length of the barcode expressed in number of characters. You can simply choose 'All'. However, the advantage of entering the total length is that the system checks again whether the barcode considered is the right one.

Field: This entry determines the field into which the barcode information should be copied.

Start: If you want to accommodate the first 'analysis unit' of the barcode, then use '1' for 'Start'. The analysis units that follow have starting values that correspond to their character positions.

Character(s): By selecting 'All' you transfer the entire barcode information to the field specified. By choosing a value between '1' and '20' you determine the number of characters entered in the field in question.

Please do not forget to save your settings.

➡ Beside text and numerical fields you can also address date fields. However you must use a 6-digit or 8-digit format for the date: it must be all numeric without separating characters of any sort. For example, '040130' would work and '20040130', too, but '30.01.2004' would not.

17 OCR definitions

OCR definitions @ archivista

Name of the OCR page definition:

Language 1:

Language 2:

Language 3:

Language 4:

Language 5:

Quality of the pages:

Check text for the right orientation: ☒

Clean page before starting recognition: ☐

Suppress scaling to minimal 300dpi: ☒

Do not use black and white conversion: ☒

Use lines for defining table cells: ☐

Table without overlapping cells: ☐

Cells with exactly one row: ☐

The form '**OCR settings**' allows you to make the settings required for internal OCR recognition. You can configure up to 20 sets of definitions which may be allocated individually. The buttons at the bottom of the form help you to jump from one definition to the next or to delete an OCR definition.

17.1 Name of OCR page definition

Each OCR definition must wear a name which you specify in line 'Name of OCR page definition'. This name will subsequently appear in submenu 'OCR settings current page'.

17.2 Languages of definition

Each definition may contain up to five languages. There is a minimum of one language. The specification of several languages helps to maximize the number of recognized symbols. If, for example, 'German' and 'French' are entered, the umlauts of both languages are valid typefaces but an Italian umlaut will not be rendered as such.

17.3 Text quality of pages

Normally, you will use the option '**Typographical**'. If you have texts which stem from a matrix printer or a typewriter the other two options help you to achieve better quality.

17.4 Further preparatory settings

'Check text orientation' means that pages that were scanned in the wrong direction are rotated so that OCR is possible. 'Clean page (e.g. delete background pattern)' deletes screen dots that otherwise would get in the way of text recognition. The option 'Suppress minimal scaling to 300dpi' prevents pages that exist in a weaker resolution from being extrapolated to a minimum of 300dpi. 'Process without black and white conversion' means that pages in gray scales or color are not converted but recognized as such.

17.5 Options for table recognition

If you have pages or documents which contain tables these options help you to make the most of them. They are especially valuable

in cases where you want to export the information contained in the tables to another Windows application.

The option **'Define cells of table according to lines'** 'tells' the OCR software that only cells with borders should be recognized. If the internal OCR does not recognize a table properly just take the original page, draw lines where you want to have them, scan the page again and have it recognized once more. In 99.9 % of all cases this will do the trick.

The option **'Tables does not contain overlapping cells'** ensures that no cell extends over another. If your table shows titles that go over more than one column this option should be turned off.

The option **'Cells contain exactly one row'** means that all cells are recognized as having one line although borders or spacing may be missing.

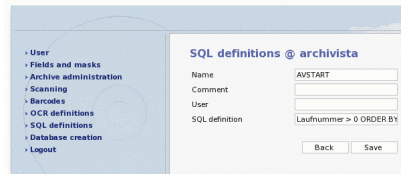
17.6 Time for optical character recognition (OCR)



The screenshot shows a web interface for configuring OCR settings. On the left is a sidebar menu with options: User, Fields and masks, Archive administration, Scanning, Barcodes, Form recognition, OCR definitions, and Time for optical character recognition (OCR). The main content area is titled 'Time for optical character recognition (OCR) @ archivista'. It contains a checkbox for 'Limit optical character recognition temporally', and two input fields for 'Starting point for recognition (0-23)' and 'Ending point for recognition (0-23)'. At the bottom right are 'Back' and 'Save' buttons.

If you want the OCR to run in the background only at specific times and not continuously – for example to avoid that the box' other processes are slowed down – you can limit the period during which OCR is active. Just fill in the appropriate starting and end time in hours (0-23).

18 SQL definitions



The SQL definitions described here serve the purpose of providing users with specific record selections at start-up. The name of the SQL definition must be AVSTART in all cases.

The individual fragments must be separated by a space each and are to be used in the sequential order described below.

```
SELECT * FROM Table WHERE (Conditions)  
ORDER BY (Sort)
```

➡ The underlined part of the SQL fragment must not be entered. It will be added automatically when the system carries out the initial selection.

(Conditions): The conditions serve to restrict the selection of records. You can use the ('%') sign as wildcard symbol, however, you must use it according to SQL standard. Leaving away the condition 'WHERE' would result in all records being selected.

(Sort): After the fragment 'ORDER BY' the field names follow according to which the sorting takes place. You can use several fields. The abbreviations 'ASC' and 'DESC' result in ascending and descending sorting, respectively. 'ASC' is optional, i.e. when nothing is mentioned an ascending sorting is effected. If you omit the condition 'ORDER BY', no sorting will take place.

Example:

```
Laufnummer > 0 ORDER BY Datum DESC, Laufnummer DESC
```

19 Form Recognition (ArchivistaBox)

19.1 Introduction

With form recognition it is possible to automatically process scanned pages that have the same structure or arrangement of items on the page.

➡ The prerequisite for the use of this module is that it has been purchased and activated.

Let us assume that we want to archive suppliers' invoices. The look and structure of the invoices is the same in all cases, i.e. the invoice date and the invoice number are to be found in fixed positions. We would like to extract both pieces of information from the invoices of the same type (forms) and copy them into the pertinent fields of our archive and we would like to split the pages automatically into documents.

This means that we scan all invoices as one batch, then the form recognition module splits them into individual documents, assigns automatically invoice numbers and dates, and files them in the archive.

In order that ArchivistaBox can do this, we must enter forms in `WebAdmin`. As a second step these forms must be activated in scan definitions (see 15). The actual start of the form recognition process coincides with the triggering of the scanning process in the `WebClient` (see 9.4).

The form recognition module allows us to manage several forms. These are defined in `WebAdmin` and menu item `Form recognition`. One can add individual forms as so called masks. One mask corresponds to one form definition. Within a definition there are objects telling us where on the form exactly a piece of information can be found (e.g. invoice number), to which

type it belongs (e.g. numerical field) and in which field of the archive it should be stored.

Thanks to good text recognition software the extraction of the information works well, however, we encounter the problem that scanning invariably involves certain inaccuracies. The invoice is not scanned straight, there is paper skew or, the page was printed awry in the first instance.

For all these cases there is the `logo recognition`. Thanks to it the first step is a logo search on the page and a subsequent positioning of the objects in accordance with the logo. This enables the recognition process to extract the needed information in a very exact manner no matter how much paper skew there was during scanning.

➡ Whenever possible the logo recognition should be used together with the form recognition. If this is not a possibility because the pages do not contain a logo, for example, one must add tolerance to the areas to be recognized (ca. 3 millimeters).

In the following the functions in connection with adding forms are described. Each mask definition corresponds to a form definition.

19.2 Managing forms

Form recognition allows you to manage several form definitions.

19.2.1 Selecting a definition

To activate a definition click on the definition name in question in the drop down list. Then click `Select`.

Form recognition @ archivista

Mask	hhw	Select	Delete
	hhw	Rename	
	kisling	Create	
	metabo		
	Petermann AG		
Desired ocr definition	hhw	Save	
Desired logo definition	hhw	Save	

>New	Name
>Edit >Delete	Rechnung
>Edit >Delete	Datum

19.2.2 Deleting a definition

The activated definition can be deleted by clicking Delete.

19.2.3 Renaming a definition

The activated definition can be renamed. This is effected by clicking Rename.

19.2.4 Creating a definition

If you want to add a new mask, enter its name in the field preceding the button Create and press Create.

Form recognition @ archivista

Mask	hhw	Select	Delete
	hhw	Rename	
		Create	
Desired ocr definition	hhw	Save	
Desired logo definition	hhw	Save	

>New	Name
>Edit >Delete	Rechnung
>Edit >Delete	Datum

19.2.5 Allocating OCR

For the extraction of the objects each form definition is in need of a text recognition software. That is the reason why each form recognition must be allocated an OCR definition (see 17).

➡ If there are only numbers to be extracted (e.g. invoice number and date), it makes sense that for the OCR definition only 'English' is chosen or, if available, 'numbers'.

Form recognition @ archivista

Mask	Petermann AG	Select	Delete
	hhw	Rename	
		Create	
Desired ocr definition	hhw	Save	
Desired logo definition	Multi	Save	

>New	Name
>Edit >Delete	Rechnung
>Edit >Delete	Datum

19.3 Editing objects of currently active definition

By clicking on New a new object can be created in the active definition.

Form recognition @ archivista

Object name	Date
Position left (mm)	86.0
Position top (mm)	56
Width (mm)	14.0
Height (mm)	6.0
Type	Date (/)
Length from	
Length to	
Field	Datum
Begin	
End	
Script name	
Activate test	<input checked="" type="checkbox"/>

Back Save

The following properties can be determined:

Object name

Each object carries a name. This serves for organisational purposes but it does not influence the recognition.

Position

By means of the fields `Position left` and `Position top` the distance from the left and the upper page margins may be defined. The upper left corner constitutes the point of origin. Thereby the page length corresponds to the y-axis and the page width to the x-axis.

➡ **Example:** The invoice number can be found 9cm from the left and 6cm from above. If you work with logo recognition you enter 89mm for 'Position left' and 59mm for 'Position top'. If you work without logo recognition, you must calculate approximately three millimeters of additional leeway, i.e. enter 86mm and 56mm, respectively.

Width/Height

The measures of the object can be determined with the help of the elements `Width (mm)` and `Height (mm)`.

➡ **Example:** The invoice number is exactly 1.2cm wide and 0.4cm high. In analogy to the example regarding position above one would enter 14mm for width and 6mm for height in the case of logo recognition, or 20mm for width and 12mm for height if one works without logo recognition.

Type

With this field you can determine of which type the piece of information is that you want to capture. There are the following choices:

- Text: entire text is accepted.
- Numbers: The first number in the recognized text is considered valid.
- Date (.): The first date in format d(d).m(m).yy(yy) is extracted. Examples: 13.4.07 or 20.12.2006
- Date (/): The first date in format m(m)/d(d)/yy(yy) is extracted. Examples: 04/13/07 or 12/20/2006

Length

The fields `Length from` and `Length to` enable you to restrict the length of the object. If, for example, only six digit numbers are to be recognized, you enter '6' in both cases. If six to eight digit numbers are to be accepted you enter '6' in field `Length from` and '8' in field `Length to`.

➡ If length is unknown, leave the fields empty.

Field

Element `Field` determines to which field in the archive the recognized text of the object is copied.

Begin and end

With these two options you can define which parts of the recognized text are to be copied to the previously specified field.

➡ **Example:** We have the number '12220001' and would like to extract only the first four digits. In this case enter '1' for `Begin` and '4' for `End`.

Script name

This option is used to postprocess (further improve) the recognized text. By entering a script name here you effect that during the form recognition process the mentioned script is run. As first and only parameter the script takes the file name of the recognized text. The script must open this file, process the text and display it on the console. The form recognition takes the script output and files the value in the field in question in the archive.

➡ **Example:** From a date that was recognized only the year should be extracted. The Perl script below does just that job:

```
#!/usr/bin/perl
use strict;
my $file = shift; # get the file name
open(FIN,$file); # read the file
binmode(FIN);
my @lines = <FIN>;
close(FIN);
my $txt = join("",@lines); # get the whole text in one variable
$txt =~ s/\r/ /g; # replace all return with space
$txt =~ s/\n/ /g; # " all newlines
$txt =~ s/\t/ /g; # " all tabs
$txt =~ s/\s\././g; # " space and point goes to point
$txt =~ s/\.\s/./g; # " point and space goes to point
$txt =~ s/{2}/ /g; # replace two 2 spaces with 1
$txt =~ /^(\.*) ([0-9]{2,2}) (\.) ([0-9]{2,2}) (\.) ([0-9]{4,4}) (\.)$/;
if ($2 ne "" && $4 ne "" && $6 ne "") {
    $txt="$6"; # if we got a day,month and year, give back the year
} else {
    $txt=""; # don't give back anything
}
print $txt; # print it out (give it back to the form recognition)
```

In order that the script may be run it must be stored in the path mentioned below.

/home/data/archivista/cust/formrec

To copy a script into this folder you can proceed in the same manner as when you prepare a logo for form recognition. For more information see 19.4.2.

Activate test

With this option you can switch on a test. The form recognition module creates a PDF file for each page showing you how the object in question is handed over to the OCR software. Positioning mistakes are easily traced this way.

It may happen that when the area in which the object to be recognized is specified too small or too big, an object and with it the entire form is not properly recognized.

To execute a test switch on this option. Note that only one object can be tested at a time. Then, scan a few pages with form recognition turned on. Wait until the pages have been processed. After that you can open the generated PDF file in the `WebClient` and check whether the object has been positioned correctly.

➡ **Example:** In the following form the invoice number filed in field `Title` is wrong. It should be 196625 but we see in the `WebClient` that only the number 19662 appears.

Document	Pages	Date	Archived	Publish	Title	Kontinent	Geschichte	Land	Region	Firmennummer	Firmenname	Dokumenttyp	PDF
63	1	07/11/2007	Yes		19662								Download

When you look at the PDF file (you can do this by clicking the field `Download` at the end of each line), you see by the fine gray line (dyed red for illustration purposes) what the reason may be.

Hommel Hercules Werkzeughandel		HW	
Werkzeuge und Werkzeugmaschinen			
HOMMEL HERCULES		Telefon	(06204) 739-0
HEIDELBERGER STR. 52		Telefax	06204/739222
DE - 68519 VIERENHEIM		USI-Kd.Nr.: HWW:	DE 146279926
Firma		USI-Kd.Nr. Kunde:	
HHW SCHWEIZ AG		STEUERFREI DREITLAND	
		RECHNUNG NR.	
		196625	Kd.Nr.: 80311
Bei Zahlung und Rückfragen bitte stets angeben			

When you look at the frame around the invoice number, you find out that it is situated a bit too much to the left and that it includes the line which it adjoins. Such a line may prove to be a stumbling block for the text recognition software.

19.4 Logo recognition

Logo recognition is a subcategory of form recognition. When scanning it may easily happen that there is paper skew. This in turn may cause troubles for the form recognition as it is dependent on the exact position of the objects to be recognized.

Logo recognition means that the logo previously stored in the system is recognized on the page. On the basis of the logo's position the system is enabled to decide whether there was paper skew during the scanning process and to correct this, if necessary, in order that the form may be recognized without any problems.

19.4.1 Creating a logo recognition definition


To create a new logo recognition definition you proceed as follows:
in menu `Form recognition` in `WebAdmin` you find the item `Logo recognition`. When you click it you see the following:

Logo recognition @ archivista

New		Name
Edit	Delete	hhw
Edit	Delete	kisting
Edit	Delete	metabo

By clicking **Edit** you can edit existing logo recognition definitions and by clicking **Delete** you can erase existing definitions. To create a new definition click **New**. You will see the following mask on your screen:

Logo recognition @ archivista

Definition name	Petermann
Logo (image)	 petermann.png
Position of logo in mm on sheet (X)	25.0
Position of Logo in mm on sheet (Y)	25.0
Number of pixels (step 1)	
Number of contours (step 2)	
Tolerance (step 2)	
Image reduction (step 2)	
Max. rotation of logo (step 2)	
Rotation iteration (step 2)	
Accuracy within logo recognition (step 3)	0.7

Back Save

In the following the individual functions are described:

- **Name of definition**

With this field you give the definition a name. In our example it is Petermann AG.

- Logo (image)

In order that the logo may be recognized it must be stored as image file on ArchivistaBox. Fill in the file name here.

- **Position of logo in mm on sheet (X) and Position of logo in mm on sheet (Y)** Imagine the form as the 4th quadrant of a Cartesian coordinate system in order that the page width constitutes the x-axis and the page height the y-axis. The origin, i.e. the point from where measurements are taken, is in the left upperhand corner where the two axes intersect. The horizontal and vertical distances of the logo from this point on the axis are the values that must be entered in these fields.

- **Number of pixels**

- **Number of contours**

- **Tolerance**

- **Image reduction**
- **Maximum rotation of logo**
- **Rotation iteration**
- **Accuracy within logo recognition**

As mentioned in 19 an exact definition of the area where the object is to be found is absolutely necessary. In the same vein it is utopian to assume that it is possible to scan a document without it being slightly skew. This is the reason why we must give the system some tolerance. This field contains the accuracy factor whereby '1' means 100 percent accuracy. In the illustration above you see the factor 0.7 as a workable average.

➡ The options **Number of pixels**, **Number of contours**, **Tolerance** and **Image reduction** cannot be used yet.

After you have filled in the necessary fields you must save your settings.

19.4.2 Creation of a new logo

Firstly, a page containing the logo must be scanned. In order that good results are achieved later during recognition you should take careful note of the following points:

- Scan letterhead page in color at 300dpi.
- Set quality to 90 at 14.3.1.
- Use a high-quality copy or original of the letterhead.

The document containing the logo must be exported to a USB stick. See 31.8 for a description of the export to a USB stick.

On the stick you find the exported file in folder 'transfer' or 'exchange'. Open it with an image editing software (e.g. Gimp). Cut the logo out carefully and align it horizontally.

Hommel Hercules WRONG!!!!

Hommel Hercules RIGHT!!!!

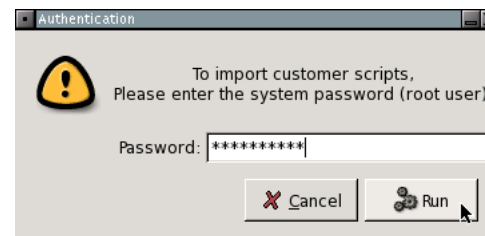
➡ Note: It is exceedingly difficult to scan a page in a way that it is absolutely straight. Nevertheless, the logo must be extremely straight because otherwise the inaccuracy produced during logo recognition regarding the rotation angle can be considerable. This in turn leads to wrong positioning of the objects. **Hence, adjust the logo so that it is in an absolutely horizontal position!**

Then, save the file in PNG format. Please do not use the JPG format because the quality of the logo would suffer.

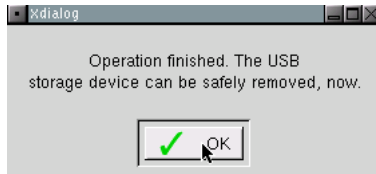
Then, take an empty (but formatted) USB stick and create a folder named 'cust'. Attach the USB stick to the ArchivistaBox. Confirm export of the 'customized' scripts.

Open the USB stick with another computer. If you do not find a directory 'formrec' inside the 'cust' folder, create such a folder (i.e. 'formrec' in directory 'cust').

Copy the logo file into the new folder 'formrec'. Copy the files back to the USB stick. ArchivistaBox recognizes the folder cust/formrec automatically. The following dialog box will appear on your screen:



Enter the password and click **RUN**



You receive confirmation that the process has been successfully completed. Click **OK**.

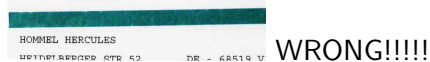
Now the logo in question is saved on your ArchivistaBox. In order that the logo recognition can take it as its basis you must not forget to enter the file name under menu item *Logo recognition*.

➡ Attention: In order that you do not overwrite existing logos on your ArchivistaBox you must export the 'cust' folder before you add the new logo to it and copy it back to ArchivistaBox.

19.4.3 Problematic logos from pre-printed letterheads

When you use pre-printed letterheads you must be very careful!

Hommel Hercules Werkze



Hommel Hercules Werkzeuge



The first picture shows you part of a pre-printed letterhead. All that is green is part of the pre-printed logo. What is printed in black was printed onto the sheet afterwards. As you may easily notice the printing is awry.

If you use the green bar as logo, the objects cannot be properly positioned because the information was not printed in alignment with the logo (here: bar). In such a case use that part of the letterhead that was printed later (here: 'Hommel...'); the objects may be awry but they are correctly awry in relation to the logo. The form recognition module will not have any problems to extract the objects properly.

➡ Note: It goes without saying that this procedure is only possible if there is an element in the individually printed part of the letterhead that lends itself to logo recognition. Printed script nameplates of a font size of 15pt or more as well as any frames or lines make good 'logos' in our sense of the word.

19.5 Summary

Below there is a summary of the individual steps in order that it may be easier for you to come to grips with the wealth of information in this chapter:

1. Scan form with logo
2. Export and edit file, save logo on ArchivistaBox
3. Turn on **logo recognition**
4. Define mask in **form recognition**.
5. Define object(s) in **form recognition**.
6. Set **scan definition** for form recognition. Important: 300dpi!!!
7. In general, when setting up form and logo recognition it is best to stick to absolute precision.

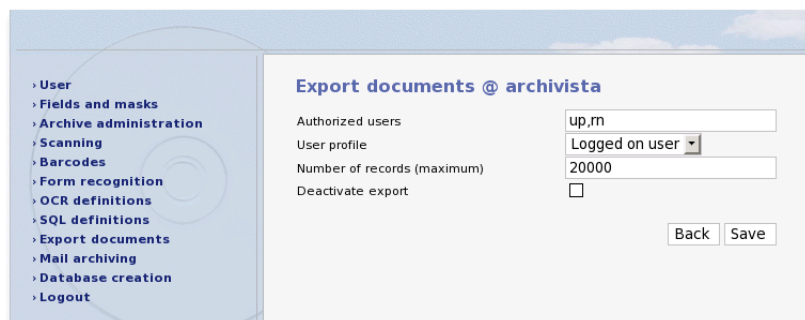
20 Exporting documents

20.1 Introduction

You can use ArchivistaBox to publish existing archives at any time. This is described in detail under 28. If you only want to export specific files, certain users that can create a database from a current query in the WebClient can be specified in this form. A new (copied) database created in this way can then be published.

This is useful, for example, if you provide scanning services and you want to deliver scanned documents as self-supporting archives. Alternatively, if you are a custodian, you may want to make the collected documents available to specific clients. You may also want to export an annual archive. In all of these cases, you can use the following form to make the relevant settings:

20.2 Exporting files in WebAdmin

The screenshot shows a web interface for 'Export documents @ archivista'. On the left is a sidebar menu with options: User, Fields and masks, Archive administration, Scanning, Barcodes, Form recognition, OCR definitions, SQL definitions, Export documents (highlighted), Mail archiving, Database creation, and Logout. The main content area has the title 'Export documents @ archivista' and four input fields: 'Authorized users' with the value 'up,m', 'User profile' with a dropdown menu showing 'Logged on user', 'Number of records (maximum)' with the value '20000', and 'Deactivate export' with an unchecked checkbox. At the bottom right are 'Back' and 'Save' buttons.

For 'Authorised users', enter the names of all users that are authorised to call this function in the WebClient as well as the Rich-Client.

For 'User profile', you can choose whether the user profile of the user that is currently logged in or the user profile of a different user

is to be utilised. For example, this allows you to export a database with a 'guest' user. In other words, the exported database can only be accessed using the 'guest' user account.

For 'Number of records (maximum)', you can determine the number of data records that can be exported. '0' means any number and positive numbers indicate the maximum number of files that can be exported.

You can use the 'Deactivate export' function to block the export temporarily.

20.3 Exporting files in the WebClient

After you activate the export, you can login with a previously defined user name. You can then find the menu item 'Export documents' in the WebClient under 'Choose action'. When you select this function, the system asks you to specify the required database name. When you enter a name and confirm the action, an export is carried out with the data records that are currently selected.

21 Mail archiving

21.1 Introduction

You can use mail archiving to archive mail messages (Emails). You have to perform the following steps in order to do this:

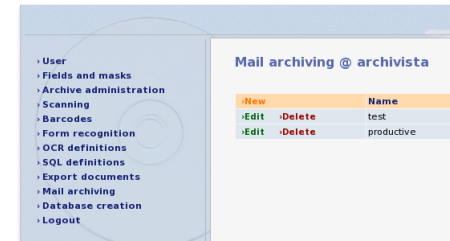
- Mail server with IMAP log (including SSL support).
- Mail archiving must be activated in the system menu (see 25.12.1 for more information).
- Mail archiving must be set up in WebAdmin for each database.
- Manual or automatic mail archiving must be completed.

This chapter only discusses the third point, that is, setting up individual mail folders that are to be archived in a database.

Before going into the individual settings options, it is necessary to point out that the following events occur during mail archiving:

- A mail message is picked up by the server.
- The mail message is analysed.
- The message itself is archived with the option of archiving all PDF files and images it contains.
- An Archivista file is created for each mail message. The 'Subject', 'From', 'To' and 'Cc' fields as well as the date can be indexed automatically.
- The message itself is zipped and is also saved.

21.2 Individual definitions



If you select the menu item 'Mail archiving', the system displays all definitions for mail archiving that are active in the database you activated in WebAdmin.

If no definition exists yet, you can create one with '**New**'. You can use '**Edit**' and '**Delete**' to edit the definitions.

21.3 Editing a definition

The screenshot shows the 'Mail archiving @ archivista' web interface with the 'Edit definition' form. The form has a sidebar menu on the left and a main area with the title 'Mail archiving @ archivista'. The form fields are as follows:

Field	Value
Definition	saved
Mail server (host)	smtp.archivista.ch
Port for connection	143
User account	mailaccount
Password	*****
Encryption (SSL)	<input type="checkbox"/>
Folder for archiving (xx.yy)	INBOX.GESICHERT
Field for sender (from)	MailFrom
Field for additional receivers (cc)	MailCC
Field for receivers (to)	MailTo
Field for subject	Titel
Owner	
Minimal age in days	
Remove mails	<input type="checkbox"/>
Move mail to folder	
Mail restore to folder	
Definition for archiving process	A4 (1 Bit - BW)
Do not archive attachment	<input type="checkbox"/>
Not activated	<input checked="" type="checkbox"/>

At the bottom right of the form are 'Back' and 'Save' buttons.

You then have to enter a name for the '**Definition**' (this is 'saved' in our example). You should not change this name after this. Oth-

erwise it may not be possible to copy the archived mail messages back into the appropriate folder.

For **'Mail server (Host)'**, **'Port for connection'**, **'User account'**, **'Password'** and **'Encryption (SSL)'**, enter the connection data for the mail server.

In **'Folder for archiving (xx,yy)'**, you can specify multiple folders that you want to archive. Use commas to separate the entries. You cannot use any other characters to separate them.

In **'Field for sender (from)'**, **'Field for additional receivers (cc)'**, **'Field for receiver'** and **'Subject'**, you determine fields in the database that are to be used for indexing the mail messages when they are archived.

In **'Owner'**, you can specify that the archived mail messages can always be assigned to a specific user or group.

You can use **'Minimum age (in days)'** to ensure that mail messages are only archived after a certain period of time.

In future, you may be able to use **'Remove mails'** to delete mail messages after archiving. *This option has not been activated yet.*

You can use **'Move mails to folder'** to specify the folder to which a message is to be sent after archiving. This allows you to ensure that mail messages are not archived more than once.

You can use **'Restore mails to folder'** to specify the folder to which an archived message is to be returned from the WebClient when you click on the 'Mail' option in 'Download'.

You can use **'Definition for archiving process'** to use a scan definition for the post processing of mail messages. Here, you also specify whether mail messages are to be archived in black and white, grey scale or colour.

You can use **'Do not archive attachments'** to ensure that mail attachments are not added as image pages of the actual message.

In each case, the whole message is archived so that it (and the attachments) can be restored at a later stage.

You can use **'Not activated'** to specify that a definition is temporarily inactive.

21.4 Possible problems during setup

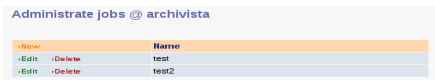
If at first you are unable to activate mail archiving, please refer to the log files where important clues can be found.

You must also take into account the fact that mail archiving will only work if you activate it beforehand in the system menu under 'Mail archiving'.

22 Administrating jobs

22.1 Introduction

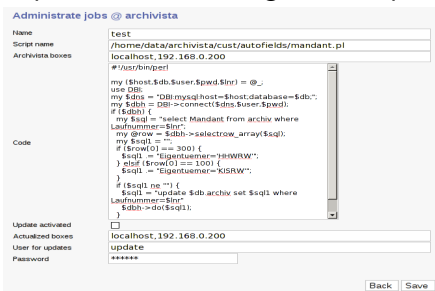
Once you are working with more than one ArchivistaBox, you must transfer the programs which are used during the execution of a processing job (e.g. scanning run) to each ArchivistaBox. So that you can do this centrally, you will find the 'Administrate jobs' menu point under WebAdmin.



➡ A prerequisite for the use of this module is that it has been purchased and activated.

22.2 Example

Per program which you want to manage, you must create an entry under 'Administrate jobs'. We would like to present the most important values using an example:



For clear identification use 'Name'. For 'Script Name', use the entire path including the program. For 'ArchivistaBoxes', enter the IP addresses of the monitored Archivista Boxes. For 'Code',

we find the program code. If 'Update is activated', then the corresponding programs are also monitored in the local directories of each ArchivistaBox. Under 'Updated ArchivistaBoxes', you will find the IP addresses of the ArchivistaBoxes which already possess the current version of the program.

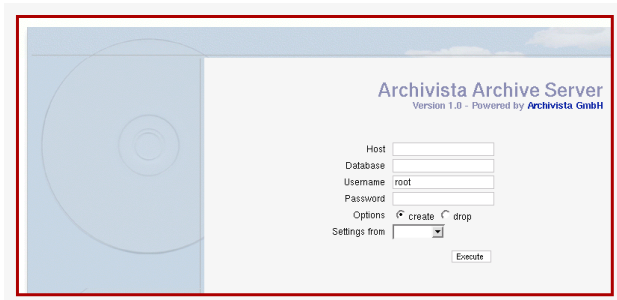
In 'User for updates' and 'Password', you must specify the user which you previously set for the same ArchivistaBox (Master) under 25.14.1.9. The last two entries are necessary in order for the monitored ArchivistaBoxes to be able to inform the Master Instance whether or not the current version is available on the peripheral station. This requires more rights than a normal user would normally receive; therefore an appropriate user must be created under 25.14.1.9.

➡ When a document is saved, there is always a check if there is a script with the same name as the database with which the user is working. This script must be manually transferred to ArchivistaBox. For the database 'archivista' it would have to be copied to the following location:

/home/data/archivista/cust/autofields/archivista.pl

Please note that the script needs rights for execution (e.g. for the example above `chmod a+x archivista.pl`).

23 Database creation



The screen used to create or delete archives is very similar to the login screen.

23.1 Option 'create'

If you want to create a new database, follow the steps below.

Host: enter the host name that you plan to use for Archivista 2014/IX.

Under **Database** you can enter a name of your choice: it must be in one word and the initial character must not be a number.

You must enter 'root' in the **User** field if you want to create a new archive.

The field **Password** must carry the root password.

The field **Settings from** gives you the possibility to 'hand down' the settings from an existing database to the new one. Select the existing database from the combo box. If you want to create a database from scratch, then you leave this field empty.

➡ Please note that you cannot delete the very first database 'archivista'.

After you click the button 'Execute' the database is created.

➡ **Important!** Please note that you can access **a new, empty** database only with user 'SYSOP'. The password is the same as

for user 'Admin'. User 'Admin' must be created first before you can use it for the new database. Note too that you cannot work with the WebClient with user SYSOP. Hence, there is no getting round creating user 'Admin' in WebAdmin before you can enter data in the WebClient.

23.2 Option 'drop'

If you want to delete a database, then click 'drop' under **Options**. Take care to enter the right names and to use the proper spellings when you enter data in the fields **Host** and **Database**. You cannot undo the process of deletion. Should you delete the wrong database by accident, then the only possible solution is to transfer the data from the backup medium.

Under **User** enter 'SYSOP' and the field **Password** must contain the right password.

By clicking the button 'Execute' you delete the database.

Part V

WebConfig

24 Administration with WebConfig

24.1 Introduction

The application WebConfig serves to configure the webbased ArchivistaBox system settings.

➡ All ArchivistaBox systems (with the exception of scan station Albis) run in RAM mode. Modifications of the system settings are initially stored on the internal disk, however, this means in the memory (RAM). The settings will be definitely stored on the hard disk only after a re-start. General information on the RAM mode can be found at 4.5.

24.2 Login

For login you have to use the root password.

[WebClient](#) - [WebERP](#) - [WebAdmin](#) - [WebConfig](#) - [Manual](#) - [Handbuch](#)



➡ Please note that the application may not be available. It may be either switched off, or the database may be busy (e.g. due

to backups). During this time it is not possible to work with Archivista WebConfig.

24.2.1 System settings

Here all settings for keyboard, network and date, time and time zones can be made.

Keyboard layout and language

Keyboard	Swiss German ▾ Update settings now
Language	Deutsch ▾ Update settings now

Network settings

Get network address (IP) by DHCP server	<input checked="" type="checkbox"/> Update settings now
---	--

Date and time

Area	Africa ▾
Time zone	
Date and time	05 : 01 : 2009 06 : 44 Update settings now

24.2.2 Mask connecting WebClient

In this menu item you determine the default values of the Web-Client login screen.

Mask connecting WebClient

Host to connect	localhost
No other hosts	<input type="checkbox"/>
Database (login form)	archivista
No other databases	<input type="checkbox"/>
User to connect	Admin
	Update settings now

24.3 Setup scan button

Define in this menu if the scan button should be operating with the current ArchivistaBox, and if so, in which database the scanned documents should be stored.

Setup scan button

Use keypad for fast scanning	<input checked="" type="checkbox"/>
Host	<input type="text" value="localhost"/>
Database	<input type="text" value="archivista"/>
User account	<input type="text" value="Admin"/>
Enter password	<input type="password" value="*****"/>
<input type="button" value="Update settings now"/>	

24.4 Backup

Here you can define the backup. Possibilities are:

- USB hard disk
- Network
- RSYNC server

Backup to external USB harddisk

Weekly days (1=Mo..7=So, 0=Never)	<input type="text"/>
Time for backup	<input type="text"/>
<input type="button" value="Update settings now"/>	

Backup to another server (Windows/Linux)

Weekly days (1=Mo..7=So, 0=Never)	<input type="text"/>
Time for backup	<input type="text"/>
Type of backup (cifs=Windows/nfs=Linux)	<input type="text" value="cifs"/>
Server (host)	<input type="text"/>
Share for backup	<input type="text"/>
User account	<input type="text"/>
Password	<input type="password"/>
Domain (if available)	<input type="text"/>
<input type="button" value="Update settings now"/>	

Backup tape drive

Weekly days (1=Mo..7=So, 0=Never)	<input type="text"/>
Time for backup	<input type="text"/>
<input type="button" value="Update settings now"/>	

Backup to another server (RSYNC)

Weekly days (1=Mo..7=So, 0=Never)	<input type="text"/>
Time for backup	<input type="text"/>
Server (host)	<input type="text"/>
User account	<input type="text"/>
Share for backup	<input type="text"/>
<input type="button" value="Update settings now"/>	

➡ **Restore data backup:** For safety reasons backup restoration is not integrated in the application WebConfig. Instead use the System menu described in 25.8.1.3.

24.4.1 Perform now

With this menu item you can start a backup promptly.

Perform now

<input checked="" type="radio"/> Backup to external USB harddisk
<input type="radio"/> Backup tape drive
<input type="button" value="Update settings now"/>

➡ **Models Dolder, Rigi, Saentis and Pilatus:** Please note that during the backup you cannot work with ArchivistaBox. Therefore ascertain that during backups no user is working with ArchivistaBox.

➡ **Models Titlis, Eiger and Matterhorn:** In the case of models Titlis, Eiger and Matterhorn backups are always carried out on the slave box. For this reason it is guaranteed that all users can work normally on the master box.

24.4.2 Log messages

Here you will find the status messages of completed backups.

Log messages

```
/libgdraw.so"
skipping non-regular file "data/archivista/images/fontforge/fontforge-20071210/.libs
/libgdraw.la"
skipping non-regular file "data/archivista/images/fontforge/fontforge-20071210/.libs
/libfontforge.so.1"
skipping non-regular file "data/archivista/images/fontforge/fontforge-20071210/.libs
/libfontforge.so"
skipping non-regular file "data/archivista/images/fontforge/fontforge-20071210/.libs
/libfontforge.la"
Sun Jun 1 09:36:29 UTC 2008
umount: /mnt/usb: device is busy
umount: /mnt/usb: device is busy
total size is 15114034432 speedup is 1.71
sent 8858529499 bytes received 240026 bytes 11452837.14 bytes/sec

Total bytes received: 240026
Total bytes sent: 8858529499
File list transfer time: 0.000 seconds
File list generation time: 21.428 seconds
File list size: 1075227
Matched data: 0 bytes
Literal data: 8855884794 bytes
Total transferred file size: 8855884506 bytes
Total file size: 15114034432 bytes
Number of files transferred: 10653
Number of files: 57886

skipping non-regular file "data/archivista/images/wine/wine-0.9.51/tools/winegcc/winegcc++"
skipping non-regular file "data/archivista/images/wine/wine-0.9.51/tools/winegcc/winegccp"
skipping non-regular file "data/archivista/images/wine/wine-0.9.51/programs/ncopy/ncopy"
skipping non-regular file "data/archivista/images/wine/wine-0.9.51/programs/write/write"
```

24.5 Services (CUPS, FTP, remote access)

Here you can define whether virtual printing into archives is permitted (cf. also 25.10) and if it is possible to deliver files by FTP upload (cf. also 25.11). Furthermore, you can determine whether it is possible to access Archivista by remote access.

Print server for archives (CUPS)

Activate print server (CUPS)

☒

Printing allowed for range

192

.

168

.

2

:

0

:

24

Update settings now

Upload server for files (FTP)

Activate upload server (FTP)

☒

Enter password

Repeat password

Update settings now

Mail archiving

Enable mail archiving

☒

Weekly days (1=Mo..7=So, 0=manually), e.g. 2-6

0

Desired time (e.g. 2:00)

2:00

Update settings now

Perform mail archiving now

Update settings now

Remote access (console with SSH)

Activate remote access (SSH)

☒

Enable it permanently

☐

Update settings now

24.6 Unlock documents

Should there be a paper jam during scanning, it may happen that a document gets locked. This means that it is no longer possible to modify the document because the system assumes that it is still being worked on. In such a case, documents can be unlocked again.

Unlock documents

Host

localhost

Database

archivista

User account

Admin

Enter password

Desired documents (i.e. 1-10 or 4)

Update settings now

⚠ Warning: incorrect specification in 'Desired documents' may result in a queuing job not being carried out properly!

24.7 Passwords ArchivistaBox

The database passwords are directly managed from ArchivistaBox. That is why all the normal passwords can be administrated directly in WebAdmin, see also 11.

Nevertheless, there are two user accounts which cannot be stored in the databases. For one it is the 'root' account (owns all rights) and for another it is the archivista user account (ArchivistaBox runs under this account). You can define the passwords for both users here. Furthermore, all user accounts for WebAdmin and WebClient can be reset. You as Administrator will need this latter function when users have forgotten their passwords.

Change password

User account

root

Password for administrator (root)

Enter password

Repeat password

Update settings now

Reset password

User account

Update settings now

⚠ Warning: Setting these passwords incorrectly or forgetting them makes any future access to ArchivistaBox impossible. Be therefore extra careful when composing these passwords.

24.8 View log files

Log messages ArchivistaBox

createdf.pl 20090105061245 start ocr
createdf.pl 20090105060743 start ocr
createdf.pl 20090105060741 start ocr
createdf.pl 20090105060740 start ocr
createdf.pl 20090105060738 start ocr
createdf.pl 20090105060737 start ocr
createdf.pl 20090105060735 start ocr
createdf.pl 20090105060734 start ocr
createdf.pl 20090105060733 start ocr
createdf.pl 20090105060730 start ocr
createdf.pl 20090105060729 start ocr
createdf.pl 20090105060222 start ocr
|

Log messages Text recognition (OCR)

12/14/2008 4:54:34 AM - End
Archivista OCR Module, FineReader Engine unloaded
PDF page is being created... (2)
PDF page is being created... (1)
PDF layout is being created... (2)
PDF layout is being created... (1)
OCR text is being saved... (370-2)
OCR text is being saved... (370-1)
Next document for OCR: 370
Connect to: localhost;archivista;root
Archivista OCR Module, FineReader Engine started
Connect to: localhost;archivista;root
12/14/2008 4:54:28 AM - Start

104 • View log files

Archivista

24.9 Text recognition (OCR)

In this menu you can prompt the OCR function to run over specific files once.

Restart OCR for document(s)

Host	<input type="text" value="localhost"/>
Database	<input type="text" value="archivista"/>
User account	<input type="text" value="Admin"/>
Enter password	<input type="password"/>
Desired documents (i.e. 1-10 or 4)	<input type="text"/>
<input type="button" value="Update settings now"/>	

24.10 Turn off ArchivistaBox

This menu item is either used to restart ArchivistaBox or to switch it off.

Turn off ArchivistaBox

<input type="button" value="Reboot"/>
<input type="button" value="Shutdown"/>

Part VI

Desktop

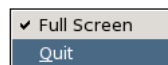
25 Desktop ArchivistaBox

If you connect a monitor to an ArchivistaBox (in the case of ArchivistaBox Dolder see 5.6.1), you will immediately get to the ArchivistaBox desktop. There (and only there) you will find the systems menu.

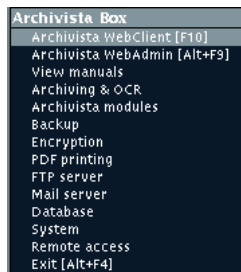
25.1 Accessing the system menu

Most system settings can be made in WebConfig (cf. 24). However, there are a few settings (e.g. backup restores) which must be done directly in the system menu of ArchivistaBox. You will need a monitor, a mouse and keyboard that must be plugged into the ArchivistaBox computer or, alternatively, you login by means of VNC (cf. 25.15.1.1).

If you work with the ArchivistaBox computer, there is normally a web application running (WebClient, WebAdmin or WebConfig). You must first get to the level of the system menu. Click the right mouse key:



Select 'Quit'. A grey screen is then displayed. Press the right mouse key again to access the system menu.



In the following the system menu functions are described one by one. Many functions require a password.

25.2 Archivista WebClient (Alt+F10)

Archivista WebClient is the application with which end users work with the archives. The WebClient is described in 6.

25.3 Archivista WebAdmin (Alt+F9)

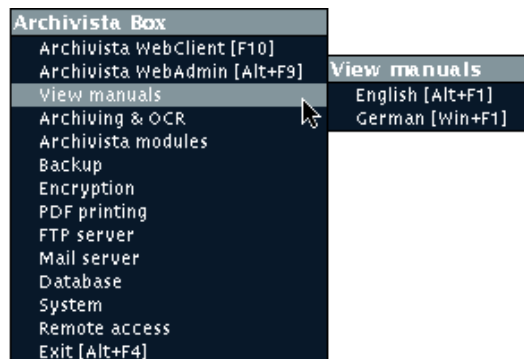
Archivista WebAdmin is the application which is used to administrate the archives (user settings, creating fields, etc.). Archivista WebAdmin is described in 10.

25.4 Archivista WebConfig (Alt+F11)

You can use Archivista WebConfig to manage the ArchivistaBox conveniently via a web application. Archivista WebConfig is described in 24.

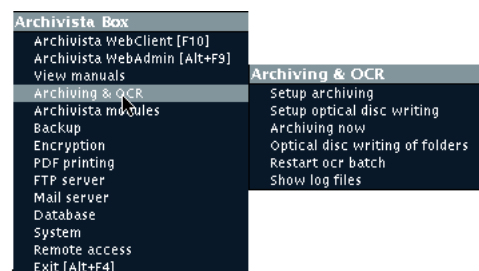
25.5 View manuals

By clicking 'View manuals' you get to the searchable PDF documents.



Use the icon 'binoculars' to search for specific expressions.

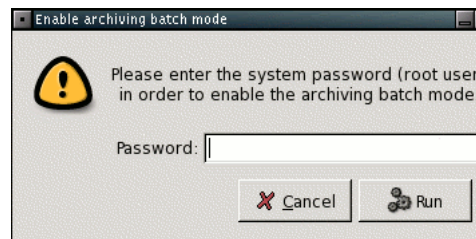
25.6 Archiving & OCR



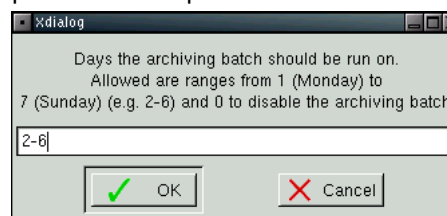
The archiving process is part and parcel of every Archivista archive. This process is supposed to run at regular intervals. It keeps the archive tidy by 'rearranging' the newly added pages and generally does everything that needs to be done in order that documents and pages may be copied to external non-rewritable media.

The options in connection with the archiving process are described in 14.3.6. This section is concerned with the actual triggering of the archiving process: when should it run and how often?

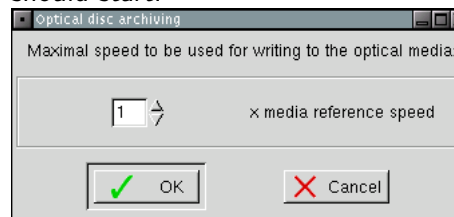
he root password must be entered first.



Then you can specify on which days the archiving process is supposed to take place.



Finally you can specify the time at which the archiving process should start.



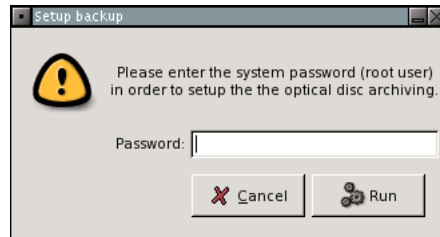
No confirmation message will be given. It is recommended to monitor the process by regularly consulting the log file 'Archiving & more' that is described in one of the following sections.

25.6.1 Setup optical disc writing

With this menu item you can configure attached CD/DVD writing devices so that later when the archiving process is run the folders in question are automatically transferred to CD/DVD.

➡ In case you would like to read an in-depth account of the Archivista archiving concept, you find the information in 30.

To arrange for automated creation of archiving folders you must get authorization first.

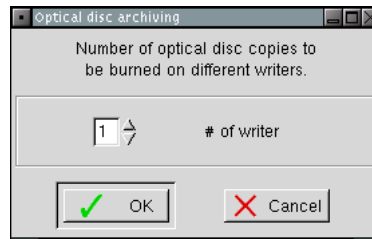


Now choose between 'CD' and 'DVD'.



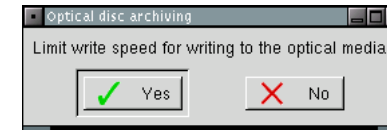
➡ The decision for or against 'CD' or 'DVD' depends on several factors relating which would exceed the scope of this manual. In general one can say that the use of CDs (in particular gold layered CDRs) is still highly recommendable.

Now choose whether you have one or two CD/DVD writers at your disposal. We unreservedly recommend the simultaneous use of two writers as with this configuration you can always create two archiving media at the same time.



Determine whether you want to give an upper limit to the writing speed. Also this is highly recommended. For one, CD/DVD writers

may not work properly at top speed, for another, the quality of the created media diminishes in proportion to the velocity with which they are written on – at least at the top end of the scale.



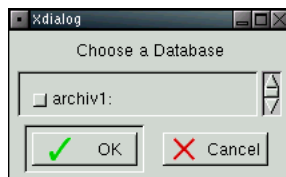
Set the desired writing speed. '1x' is admittedly too conservative for the majority of cases. When trying to find the optimum speed keep in mind the remarks above.



With this the configuration of the CD/DVD writers is finished. You can start the archiving process and wait until the first medium is written. It goes without saying that you must first put a CD/DVD medium into the CD/DVD device. And yes, depending on the number of documents the writing process may take longer. On average, a CD/DVD should be written after a couple of thousand pages.

25.6.2 Archiving now

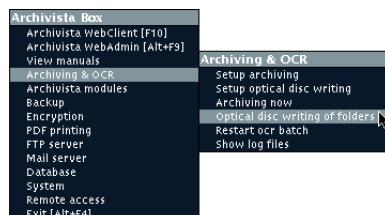
Alternatively, the archiving process can be triggered immediately. After having entered the root password you can pick the databases for which the archiving process is to be run. (The example below shows only one database.)



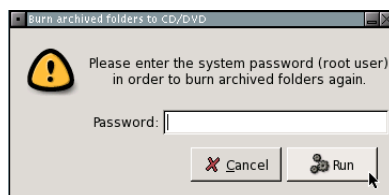
No confirmation message will be given. It is recommended to monitor the process by consulting the log file 'Archiving & more' that is described in one of the following sections.

25.6.3 Optical disc writing of folders

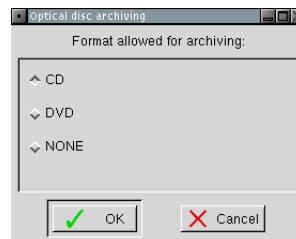
The purpose of the function **Optical disc writing of folders** is that the folders are not only saved in the database but on a non-rewritable external medium (CD/DVD). To make a copy of one or several folders proceed in the following manner:



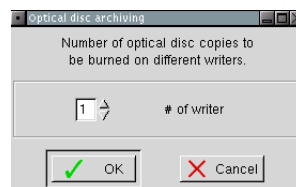
Choose function **Optical disc writing of folders** in menu item **Archiving&OCR**.



Enter the password.



Choose the medium on which you want to archive the data.



Enter the sequential numbers of the folders which you want to burn to the disk. If you specify '1-4', then the folders 1 to 4 are written to the CD/DVD. Click 'OK'.

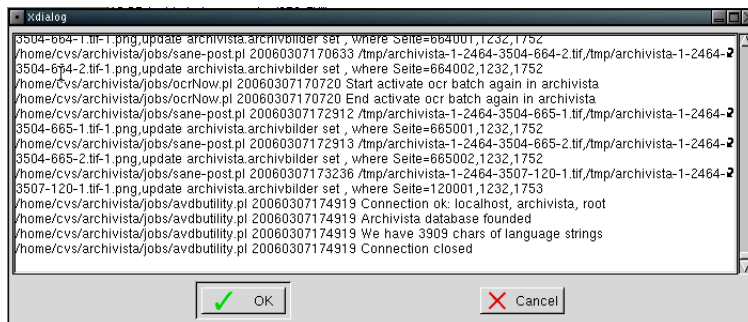
25.6.4 Restart OCR batch

This function starts the OCR as batch program. I.e. the OCR is not running continually in the background but is started manually. No confirmation message will be given. The log files described in the next section serve to monitor the process.

25.6.5 Show log file

25.6.5.1 Archiving & more

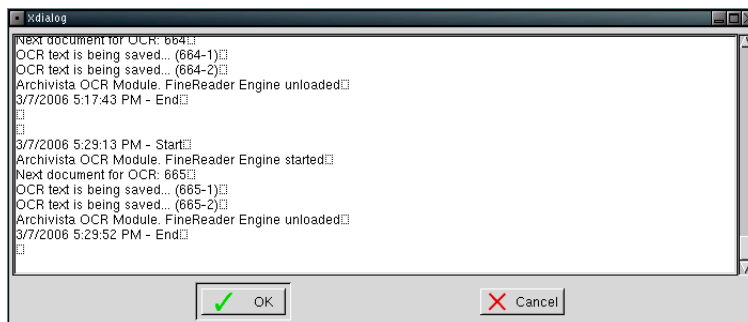
The log file gives information about whether the archiving process has been started and about the procedures already finished as well as about potential inconsistencies. There are two submenus, namely 'Archiving & more' and 'Text recognition'.



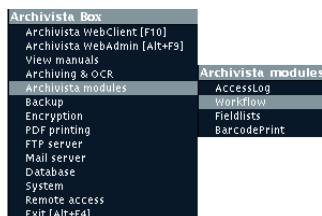
From this particular log file of 'Archiving & more' it can be seen that the OCR process was started and ended.

25.6.5.2 Text recognition

In the log file 'Text recognition' details of the text recognition process are listed.

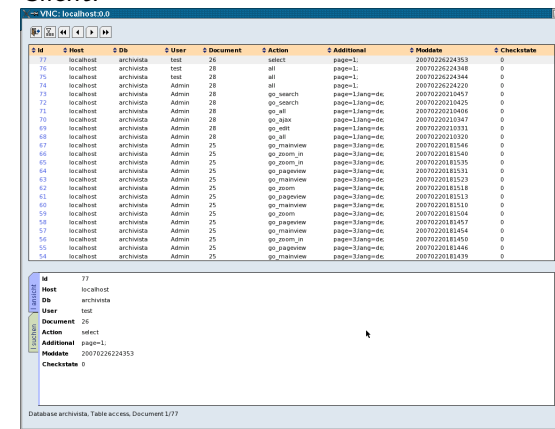


25.7 Archivista modules



25.7.1 AccessLog

This module serves to create a logfile of all accesses to the WebClient.



25.7.1.1 Turning it on

Please note that this option is normally not turned on. You must activate it as described in 14.3.10.

25.7.1.2 Slave computer

Please note that it is not possible to access an archive on the Slave Box by means of the WebClient (as all accesses can be protocolled only once). To nevertheless ascertain whether the ArchivistaBox in Slave mode does cleanly replicate the Master Box you will see, when logging into the Slave computer, an error message displaying the last modification that has taken place in the database.

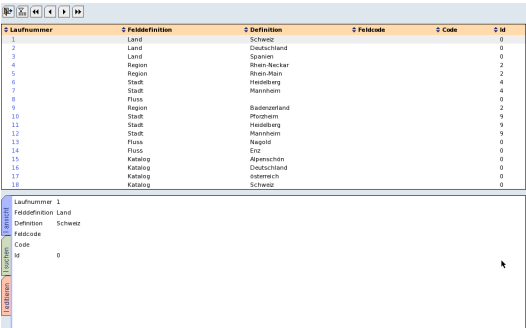
➔ The following may serve as example: Try to log into the Slave. The time appears when the last change was made. Now go back to the Master and use the WebClient to navigate from one document to the next. This causes the 'AccessLog' table to be automatically updated in the background. Now go back to the Slave and use the

Webclient to login. The generated time stamp must be different from the first because the activity registered by the 'AccessLog' table in turn influences the ArchivistaBox Slave.

25.7.2 Workflow module

The module 'Workflow' is not part of the ArchivistaBox standard package. This does not mean that there is no workflow functionality in ArchivistaDMS; it is directly integrated in the WebClient. This menu item is included here for compatibility reasons.

25.7.3 Field lists



Laufnummer	Felddefinition	Definition	Feldcode	Code	ID
1	Land	Schweiz		0	
2	Land	Deutschland		0	
3	Land	Spanien		0	
4	Region	Rhein-Neckar		2	
5	Region	Rhein-Main		2	
6	Stadt	Heidelberg		4	
7	Stadt	Mannheim		4	
8	Fluss			0	
9	Region	Baden-Württemberg		2	
10	Stadt	Heidelberg		4	
11	Stadt	Heidelberg		4	
12	Stadt	Mannheim		4	
13	Fluss	Neckar		0	
14	Fluss	Elbe		0	
15	Katalog	Abrechnung		0	
16	Katalog	Deutschland		0	
17	Katalog	Österreich		0	
18	Katalog	Schweiz		0	

Laufnummer 1
Felddefinition Land
Definition Schweiz
Feldcode
Code 0
ID

25.7.3.1 Introduction

In table 'Fieldlists' the definitions of the choice boxes are laid down. Normally, the entries belonging to the choice boxes are modified directly in the WebClient. Cf. 8.3.3. With the help of the module 'Field lists' the definitions can be modified centrally.

➡ Making modifications in table `field lists` requires considerable circumspection as links and dependencies between fields may be destroyed if one makes editing mistakes.

25.7.3.2 Navigation

There are six navigation icons:



Quit program (Alt+q)



Select all records (Alt+a)



Previous records in selection (Alt+f)



Previous record (Alt+p)



Next record (Alt+n)



Next records in selection (Alt+l)

25.7.3.3 Structure of table 'feldlisten'

Table 'field list' has the following structure:

- Field definition
- Definition
- Field code
- Code
- ID
- Sequence number

The `Sequence number` contains a sequential number.

The column `Field definition` shows those fields that were allocated either the field type `Definition` or the field type `1:N` in menu mask definition of **WebAdmin**.

Example: The field `Country` is allocated the field type `'Definition'`. Hence, `Country` will appear in column `Field definition`. The countries allocated to the field `Country` in the **WebClient** appear in column `Definition`. If we assume that the field `Country` has the sequence number `'1'` in the fieldlist, this sequence number appears also beside the field `Region` that can also be found in column `Field definition`; `Region` is of the field type `1:N` that appears in column `ID`. Explanation: the region was subordinated to the country by allocation of field type `1:N`. The entries made in field `Region` in the **WebClient** appear in column `Definition`.

Laufnummer	Felldefinition	Definition	Feldcode	Code	ID
1	Land	Schweiz			0
20	Region	Zürich			1

Fields to which the field types `Text` or `Number` have been allocated appear in column `Field code`. The number code to which a code was allocated in the **WebClient** appears in column `Code`. For more information on how to edit fields in the **WebClient** go to Chapter 8.3

Example: We are to develop a system whereby each company is assigned a number. For the purpose of illustration the name of the sample company is **Petermann AG**. We first have to create the field `Company name` and then assign it to field type `Definition`. This field appears in the `Fieldlist` in column `Field definition`. In the **WebClient** we make the entry **Petermann AG** in field `Company name`. This entry can then be found in column `Definition`. To assign a company number to this company one creates the field `Company number` and allocates the field type `Number` code to it. The codes pertaining to the companies are assigned in the **WebClient**. Subsequently they will appear in column `Code`.

Laufnummer	Felldefinition	Definition	Feldcode	Code	ID
21	Firmenname	Petermann AG	Firmennummer	2	0

The idea behind the `Multi` field is to be able to use more than one keyword of the same order for one document. The field type `Multi` is also subordinated to the field type `Definition`. Hence, each `Multi` field must be linked to a `Definition` field. Compare Chapter 13.3.6.

In the case of a `Multi` field the field with the definition type `Definition` appears in column `Field definition` and the field with definition type `Multi` in column `Definition`.

Laufnummer	Felldefinition	Definition	Feldcode	Code	ID
------------	----------------	------------	----------	------	----

In this example `'catalog'` was given the field type `Definition`. For the multi fields `catalog1` and `catalog2` Germany, Switzerland and Austria were added as entries.

subsubsectionThe mask

In the lower part of the screen there is a mask with three different tabs:

- View
- Search
- Edit

`View` shows the information that is contained in the fieldlist. The `search` mode enables the search for strings or a group of strings on the basis of individual characteristics. For example, if one wanted to know which records are linked to the record with sequence number `'3'` one enters `'3'` in `ID` in the search mask and subsequently the system returns all records subordinated to record `'3'` by means of field type `1:N`.

Behind tab `Edit` there is a window in the upper left corner that gives us the options to add new strings or to change existing strings. In addition, existing information can be

changed by first clicking on 'Edit' and then on 'Update'.



25.7.4 BarcodePrint

The module 'BarcodePrint' is not active for all customers, i.e. it is not preconfigured on the OpenSource CD and ready to use. As an alternative have a look at the RichClient application Bar-codePrint in section 16.

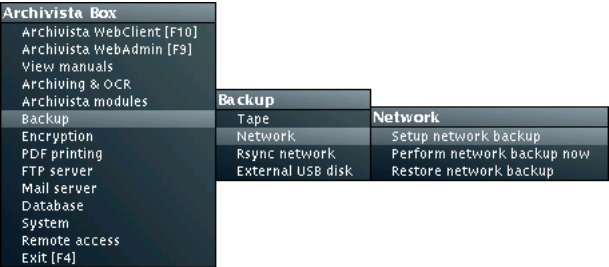
25.7.5 ArchivistaERP

These options are described in detail in 27.

25.8 Backup

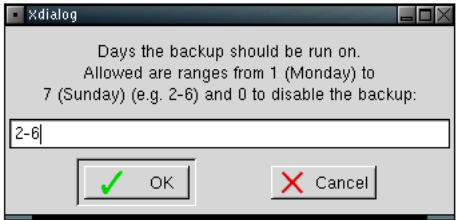
The backup function is needed to copy your data. There are a variety of ways to do backups: They are described below. During the first copy process to the backup medium all data is transmitted, during the following operations only the differences are replicated. Depending on the data volume, the first backup may take some time while the differential backups should take only a few minutes.

25.8.1 Network

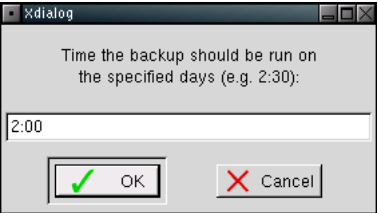


25.8.1.1 Configuring backup

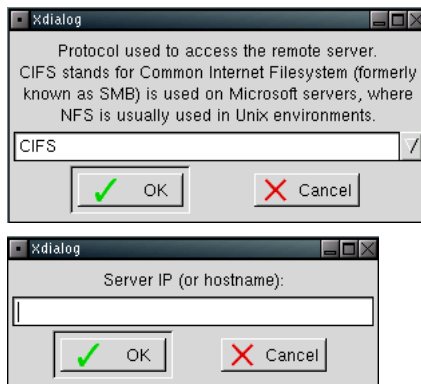
When setting up the backup you are first prompted for the root password. Then you will receive the following query:



Enter the days when you want a backup to take place.



Enter the time of day when the backup should take place. The default value '2:00' means 2 o'clock in the morning. Next you must enter the protocol used and the IP characteristics of the target computer. Entries have to be made in the following dialog boxes.



In the above dialog box enter the IP address of the computer on which the backup copy is to be saved.

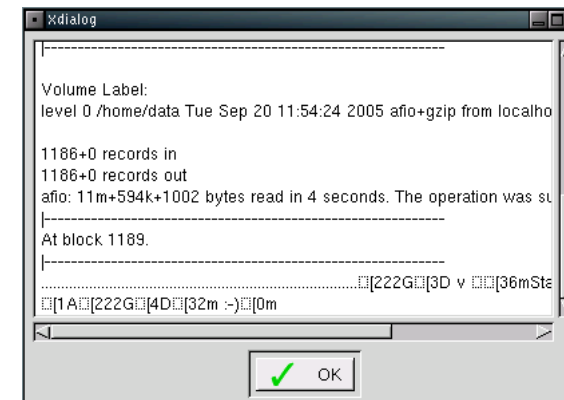
➡ There is no further confirmation. After the time of day has been entered the backup is properly set up. Each time the backup process has successfully ended (usually each morning) you will get a message to that effect.

25.8.1.2 Perform backup now

The menu item 'Perform backup now' kicks off an immediate backup-uping procedure. Click 'Perform backup now' and enter the root password: The copying process starts if a storage medium is connected.

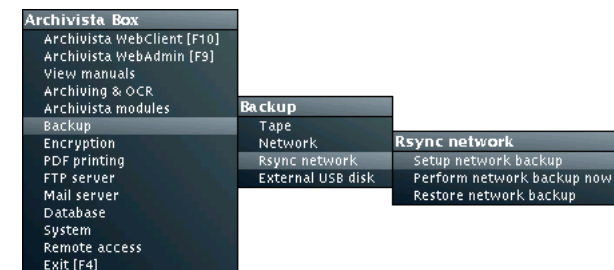
25.8.1.3 Restore backup

Running the command 'Restore backup' and entering the root password enables you to copy what is on the backup back to the hard disk of the ArchivistaBox. The more data there are the longer the process takes.



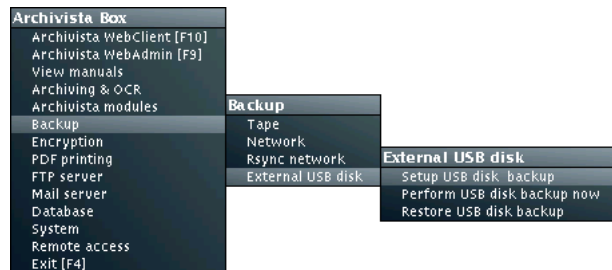
After the restoring process is finished a message pops up telling you that the database server was started again.

25.8.2 Network (rsync)



With this menu item you can also perform a network backup. However, no drive (share) will be made available, but rather a RSYNC server. This server is able to filter out the difference between the original on the ArchivistaBox and the copy which the RSYNC server has previously created more efficiently than is the case with a normal network backup.

25.8.3 External USB disk



The commands in this menu are analogous to those for a data backup to a network, except that the copy is created on an external USB hard drive. It is advantageous to write to several of these external hard drives and to store them in different places, so that in the event of a lost drive a different drive may be used.

Since common USB drives are becoming ever larger, you may also redefine how many old backups should be kept on the backup drive before a backup is definitely deleted.

➡ Please note that the drives must be set up in the 'ext3' format if any of the tables to be backed up are larger than 4 GB. You CAN use drives that are NTFS-formatted, however, the function 'Restore-on-the-Fly' is not available (cf. 25.8.3.1). Drives in ext3-format can be created with the function 25.14.1.6. Alternatively, you can use a standard Linux operating system.

25.8.3.1 Restore-on-the-Fly

If you save the ArchivistaBox ISO file in `/var/lib/vz/template/iso` under the name of `dmsbackup.iso`, then the ArchivistaBox installation is always backed up, too. In this way, you can restore ArchivistaBox from scratch from your backup.

This means that you can test any of your backups with (almost) any PC by booting from the external disk. Archivista will be set up

in the RAM so that you can start stored archives (in read-only mode!) and test it at will. Restore-on-the-fly is also very helpful in those cases when only individual documents need to be restored.

25.9 Encryption

When you install the ArchivistaBox computer in the 64-bit version, then the encryption function is already active, i.e. all data is transmitted using the HTTPS protocol, meaning that you do not need to enable HTTPS.

25.9.1 Enable https

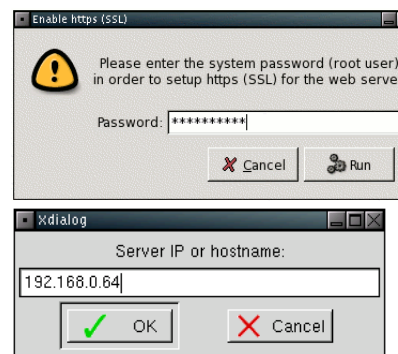
25.9.1.1 What is HTTPS?

Ordinarily web pages are transmitted by HTTP (= Hypertext Transfer Protocol). The data are not encrypted.

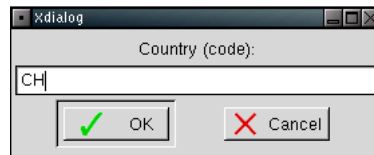
When HTTPS (= Hypertext Transfer Protocol Secure) is used the data flowing between web browser and web server are encrypted.

By running the function 'Enable https' we 'tell' ArchivistaBox that from now on it must encrypt all data that it sends to other computers in the network upon their request. HTTPS makes most sense in the case of confidential data.

25.9.1.2 Function 'Enable https' on the server side



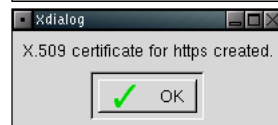
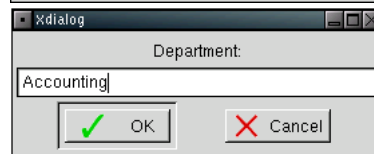
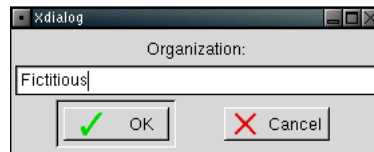
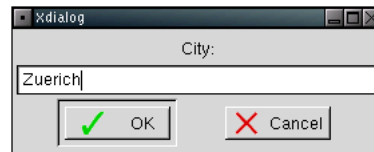
Here you must enter the IP address of the computer with which you are currently working. Remember, you are configuring HTTPS on your server just now.



The 'Country code' must be two characters long.



All further settings can be entered without any such restrictions.



25.9.2 Disable https

Here you simply enter the root password. Subsequently, the HTTPS mode is turned off again.

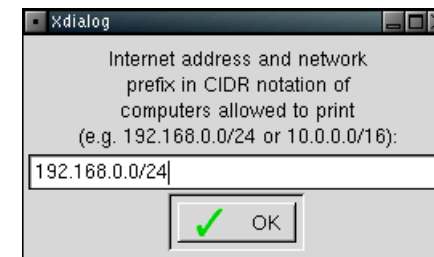
25.10 Print server

The function 'Print server' sets up a virtual printer. This enables users to print any files existing somewhere in the network to the appropriate archives.

25.10.1 Enable print server



Run the function. After entering the password you see the following message:



Enter the network subnet from which files are accepted.

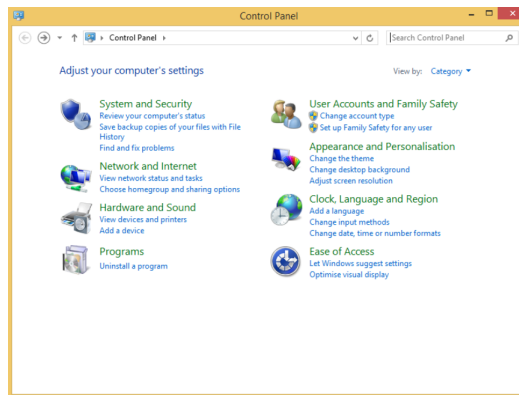
This is server-side. Please note that on client-side it is necessary to configure the Archivista printer on workstations from which documents are to be sent to ArchivistaBox.

For this purpose we need the IP address of our ArchivistaBox. By clicking 'System' and 'Display system status' we can find out this address. In our case it is 192.168.0.71.

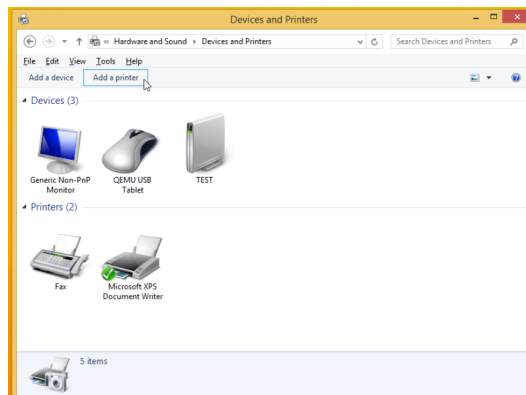
25.10.1.1 Configuring the Archivista printer

The following example stems from a German Windows 8.1 workstation. In principle, the same procedure applies to all Windows versions.

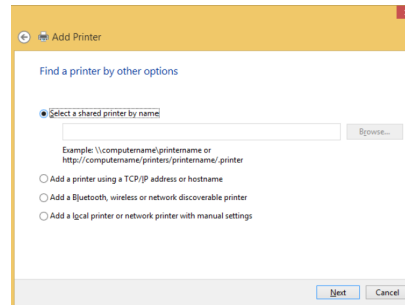
Go to 'Control Panel', then to 'Hardware and Sound' and 'View Devices and Printers'. Doubleclick 'Add a printer'. Click 'Next'.



Afterwards you specify that it is a **network** printer you want to add.



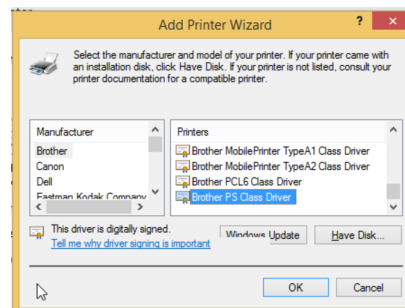
Select 'The printer that I want isn't listed'. The dialog 'Add printer' appears.



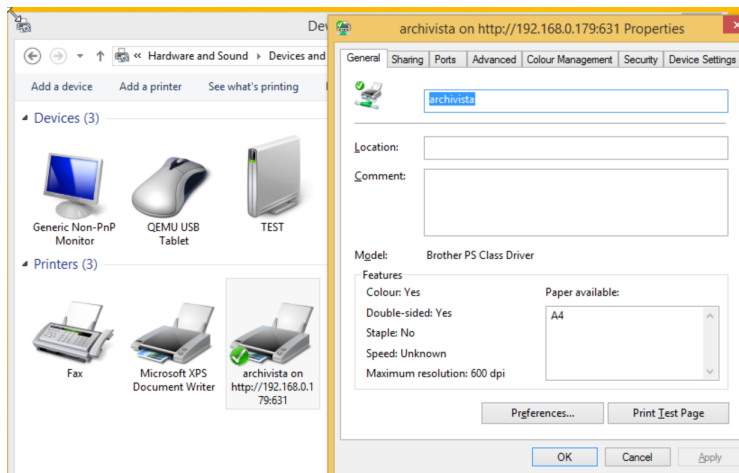
Choose 'Shared printer...' and enter the following (replacement of IP address by address of ArchivistaBox):

`http://192.168.0.100:631/printers/archivista`

Click 'Next'. You can now determine the printer type. As we work with a PS printer, we can select the Brother PS-Class-Driver printer, for example. (Simply pay attention to 'PS' appearing in the name).



If you click on 'OK' now, the printer is set up. A confirmation dialog pops up which you must confirm again. A printer named 'archivista' must now appear in the list.



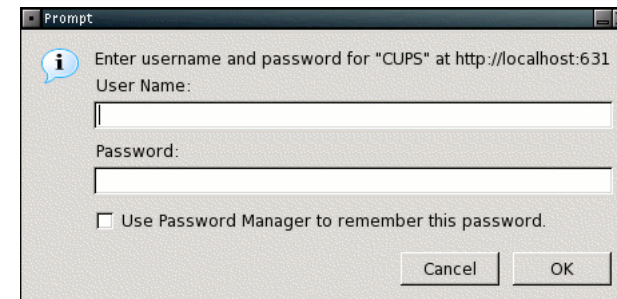
The next thing to do is to send a test page to the archivista printer. From now on printing to ArchivistaBox should work without any problems.

➡ With some Windows versions (depending on the update status these are the versions of Windows 2008, 2012 and Windows 7) the service 'Internet Printing Services' must be activated manually. If you are not successful straightaway, a Windows restart may do the trick.

25.10.2 Administration

When the print server is active, you can administrate this printer also directly from your ArchivistaBox.

After you have run the function, the following dialog box comes up and requires ID and password:



Login with user 'root' and the corresponding password. You will see a list of existing printers.

Here you can administrate the printers and create additional printing definitions – for example to print into a second database. You will need to set up separate printers for each database, i.e. if you want to print to the database 'archivehello', then you will need a printer with the name of 'archivehello'.

➡ It is often the case that you want to print certain files in colour and others in black and white in the archive. You can then set up a printer that contains the required scan definition, separated from the database name by a hyphen. For example: You can use the `archivehello-colour` printer to print in the 'archivehello' database and the system also checks whether the 'color' scan definition exists during the import. If this is the case, the options of each scan definition become valid. In other words, the print jobs are post-processed with the extended options.

25.10.3 Disable print server

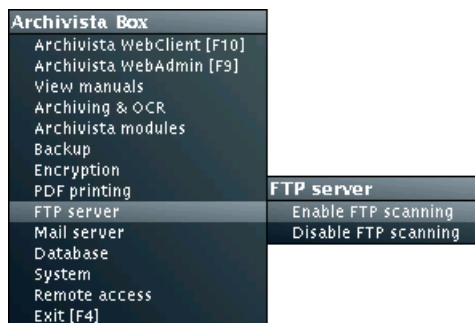
With this function you turn off the print server.

25.11 FTP Server

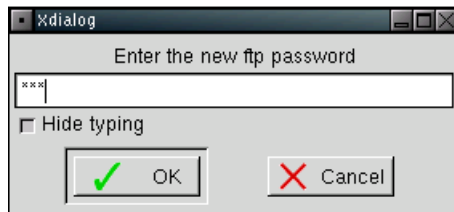
This function allows you to set up a FTP server in order that documents may be transferred to ArchivistaBox.

➡ Along with the FTP server you set up an SMB server (Windows directory). Always utilise user 'ftp' for the FTP server and user 'archivista' for the SMB directory. Both services can be accessed with the password specified here. When you access by means of SMB, it is important that you do it as workgroup user. Should access not be successful straightaway – e.g. from Windows – you can enforce the workgroup user when connecting to a network directory by prefixing the user name with the IP address and a slash: (\\192.168.0.100\\archivista).

25.11.1 Enable FTP file upload



Run the function 'Enable FTP file upload'.



As always you will first be asked for the root password. Then, you will be asked to enter a specific FTP password. You will be prompted to enter it a second time. Now (password protected) FTP access to the ArchivistaBox is granted. No confirmation message will be displayed.

|| FTP server enabled

Check 'System' and 'Display system status' to review the status of the FTP server. In the illustration above you see the entry 'FTP server enabled'. Hence, everything is in place for you to scan directly to the database on ArchivistaBox.

25.11.1.1 Upload in folder

The FTP server was originally developed to work together with FTP-capable digital copiers. However, it is now possible to send any documents to the archive via an FTP upload. The following folders exist at present:

- office: Import any files (a text extraction occurs rather than OCR recognition).
- pdf: Import scanned PDF files (OCR recognition occurs).
- tiff: Import scanned tiff multipage files (OCR recognition occurs).
- lbs: Cold import from ERP solution
- toasca: Cold import from ERP solution
- axapta: Cold import from ERP solution
- xerox: Import from Xerox copiers

➡ Please note: If you send a file directly to the subfolder (e.g. office), the file is always imported to the database **archivista**. If you want the file to be imported to a different database, you have to copy the file to the relevant subfolder (e.g. you have to send a file to `\office\archive` if you want to store it as an Office document in the 'archive' database).

➡ When importing files, the files are processed using the first scan definition. Since 2013 additional subsubfolders can be created in

the subfolders, this with the purpose of triggering the appropriate scan definition. Example: You can create a directory named 'Colour' to have documents processed with the scan definition 'Colour'. This makes sense, too, if you want to index certain fields already at this stage. Cf. 15.3.1.

25.11.1.2 Scanning by means of network scanners

There are many network scanning devices around today. The Xerox Work Centres may be mentioned as prominent examples. But also the more sophisticated copiers of other brands give this possibility. When you work with a copier you must ascertain that login takes place with the type 'FTP' or 'SMB'. Choose user 'ftp' for the FTP server and user 'archivista' for the SMB server (Windows directory). Use the previously set password.

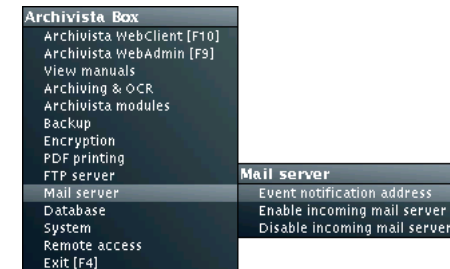
In order that ArchivistaBox knows which documents are to be processed, the scanned documents must be uploaded into specific folders. First hierarchy level: the document type, usually tiff or pdf. Second hierarchy level: the appropriate Archivista database. Two examples:

- Black and white scans for database 'archivista':
/tiff/archivista
- Colour scans for database 'archiv44': /pdf/archiv44

25.11.2 Disable FTP file upload

With this menu item you shut down FTP access.

25.12 Mail server



25.12.1 Activating mail archiving

Use this function to activate mail archiving for all databases. Please note that you then have to set up mail archiving in WebAdmin for each database. For more information about mail archiving see 21.1.

25.12.2 Start mail archiving now

You can use this menu item to start mail archiving at any time.

25.12.3 Deactivating mail archiving

You can use this menu item to deactivate mail archiving again.

25.13 Database

➡ The menu items described below serve to manage the replication of redundant ArchivistaBoxes 'Titlis', 'Eiger' and 'Matterhorn'. Since 2013 replication is fully automated. Hence, these menu items are no longer in use.

The function 'database' was needed with 'Titlis', 'Eiger' and 'Matterhorn', i.e. when two ArchivistaBoxes constitute a redundant system. One has to distinguish between the master and the slave

computer. The master computer takes the modifications (additions, changes, deletions) and transfers them automatically to the slave computer.

25.13.1 Setup master for slave

By clicking this function you make the currently active ArchivistaBox the leading one: it contains the live configuration and the live databases and any alterations that are made affect this ArchivistaBox first.

Further down there is an example of a complete master slave configuration.

25.13.2 Enable slave mode

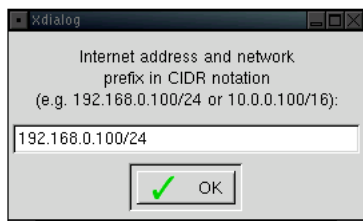
By clicking this function you make the currently active ArchivistaBox the subordinate one. The alterations made to the master box and its databases are mirrored to the subordinate box every few seconds.

The next section gives an example of a complete Master/Slave configuration.

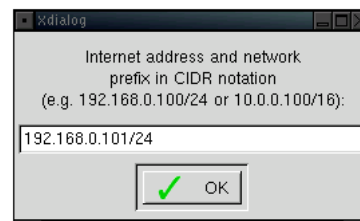
25.13.3 Example: Master/Slave configuration

25.13.3.1 Fixed IP addresses

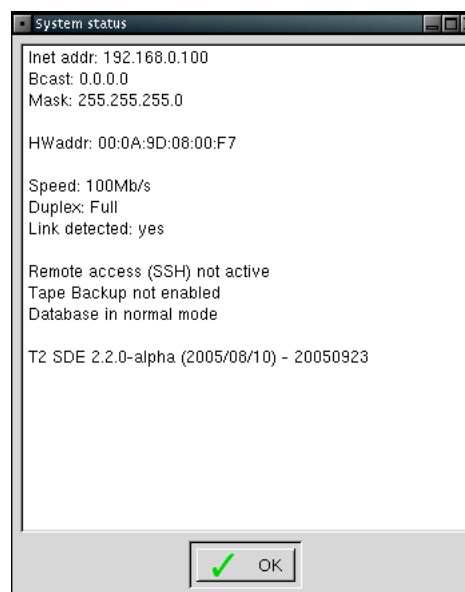
Before you can start with the Master/Slave configuration both ArchivistaBoxes must possess one fixed IP address each. Cf. 25.14.1.3.



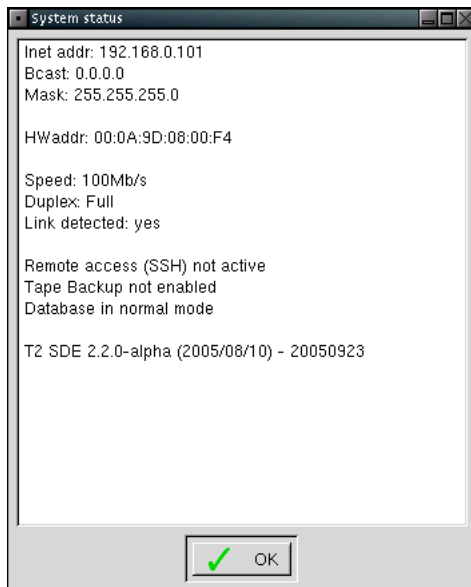
This is an example of the ArchivistaBox 'Master'.



This is an example of the ArchivistaBox 'Slave'.



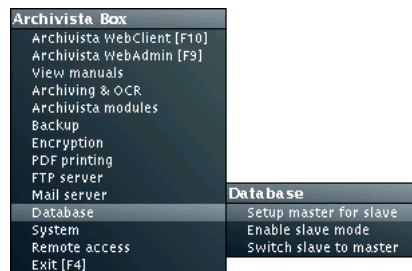
When running 'Display system status' we can see which IP address the box possesses. Above you see the 'Master'.



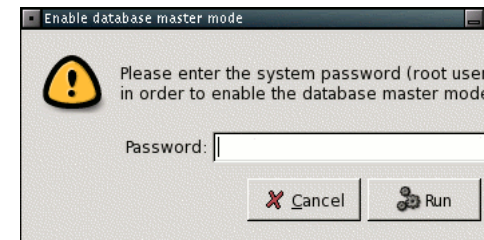
The above illustration shows the 'Slave'.

25.13.3.2 Configuring the 'Master'

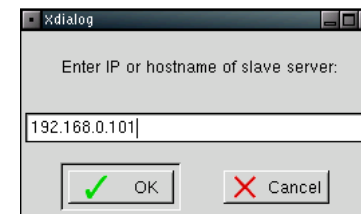
Setup the ArchivistaBox 'Master' now.



Choose the command 'Setup master for slave'.



A dialog box will ask for the password.



Afterwards you must enter the IP address of the ArchivistaBox 'Slave' (without the information after the / sign). The next steps are meant to grant the 'Slave' the right to access the 'Master'.



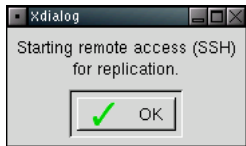
The following dialog box invites you to give the 'Slave' a name. In our example it is 'slaveuser'.



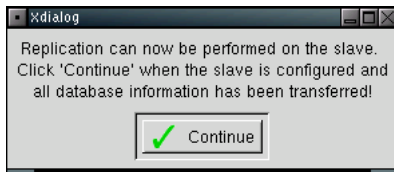
Here you allocate the 'Slave' user a password.



Unless 'remote access' was previously granted manually, the ArchivistaBox asks you now whether the keys for SSH connection should be generated. Click 'Yes'.



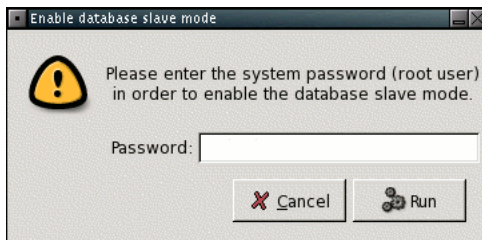
The following dialog box confirms the SSH connection. Click 'OK'.



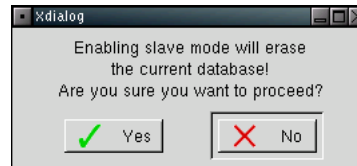
The setup of the 'Master' box is complete. Please note: Do not click 'Continue' as yet! First you must setup the 'Slave' box.

25.13.3.3 Configuring the 'Slave'

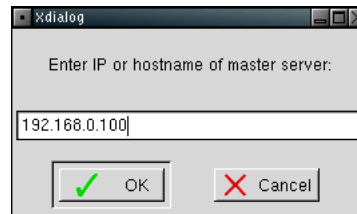
As the box intended to be the 'Slave' serves as protection of your data assets on the 'Master' and normally does not need a monitor, you can use VNC to access it.



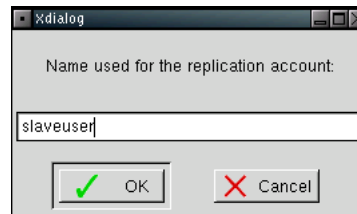
Go to the desktop of the 'Slave' and activate it by selecting menu item 'Activate Slave mode'. Again you will be asked for the password.



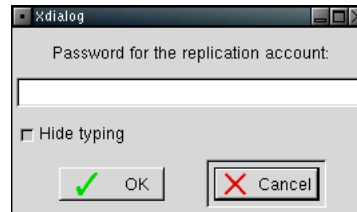
By confirming the message above you effect that existing databases are deleted from the ArchivistaBox.



Subsequently, the fixed IP address of the '**Master**' box must be entered.



Here you must enter the name of the 'Slave' user. It goes without saying that the name must correspond to the one entered in the 'Master' box. In our case this is 'slaveuser'.



The password is the same too.


```
db-slave.sh
PING 192.168.0.100 (192.168.0.100): 56 octets data
64 octets from 192.168.0.100: icmp_seq=0 ttl=64 time=0.2 ms

--- 192.168.0.100 ping statistics ---
1 packets transmitted, 1 packets received, 0% packet loss
round-trip min/avg/max = 0.2/0.2/0.2 ms
Stopping MySQL Server. .... :-)
Please enter the master server system (root user) password in order to
copy the initial database:
The authenticity of host '192.168.0.100 (192.168.0.100)' can't be established.
RSA key fingerprint is df:c8:82:e2:d4:cb:fb:f6:fc:f9:da:67:07:b9:24:6c.
Are you sure you want to continue connecting (yes/no)?
```

This message may or may not appear, but if it does, it should be answered by 'Yes'.

```
db-slave.sh
PING 192.168.0.100 (192.168.0.100): 56 octets data
64 octets from 192.168.0.100: icmp_seq=0 ttl=64 time=0.2 ms

--- 192.168.0.100 ping statistics ---
1 packets transmitted, 1 packets received, 0% packet loss
round-trip min/avg/max = 0.2/0.2/0.2 ms
Stopping MySQL Server. .... :-)
Please enter the master server system (root user) password in order to
copy the initial database:
root@192.168.0.100's password:
I
```

Then, enter the **'Master'** password.

```
db-slave.sh
The authenticity of host '192.168.0.100 (192.168.0.100)' can't be established.
RSA key fingerprint is df:c8:82:e2:d4:cb:fb:f6:fc:f9:da:67:07:b9:24:6c.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '192.168.0.100' (RSA) to the list of known hosts.
root@192.168.0.100's password:
receiving file list ... done
deleting mysql/localhost-relay-bin.005
mysql/
mysql/ib_logfile0
mysql/ibdata1
mysql/localhost-bin.011
mysql/localhost-bin.012
mysql/localhost-bin.index
mysql/localhost-relay-bin.006
mysql/localhost-relay-bin.index
mysql/localhost.err
mysql/master.info
mysql/relay-log.info

sent 42642 bytes received 2632553 bytes 28308.94 bytes/sec
total size is 612266016 speedup is 228.87
Starting MySQL Server. .... :-)
Press enter or close this window.
^[[A
```

When you see this screen the 'Slave' is ready to take up work and to replicate all data on the box 'Master'.

Please do not forget to finalise the configuration by clicking the button 'Continue' on the box 'Master'.

25.13.3.4 Testing the configuration

Out of a variety of reasons setting up such a mode is not always successful straightaway. Hence, do test your Master/Slave installation. Make a few alterations on the 'Master' and check whether they are mirrored on the 'Slave'.

If the 'Slave' inherits all alterations, no further measures are necessary. If data transfer has failed, though, you should restart both boxes. If it still does not work, you must turn the 'Slave' into the 'Master' and start from scratch. This function is described below.

25.13.4 Switch slave to master

This menu item serves to upgrade 'slave' to 'master' in case the 'master' fails one day.

After having run this function you can look in 'Display system status' what the current status is.

If the 'Slave' mode is active, then the status says 'Database in slave mode', if the 'Master' mode is active, then the status says 'Database in master mode'.

➡ If you possess a redundant ArchivistaBox solution and one box fails, it is recommended that you order a new ArchivistaBox and connect it to the upgraded 'master' box and set it up as 'slave'. During the time when you work with one ArchivistaBox only you can save the ongoing changes if need be by running an additional backup during lunchtime. See 25.8.1.2.

25.13.5 Clear master binlog

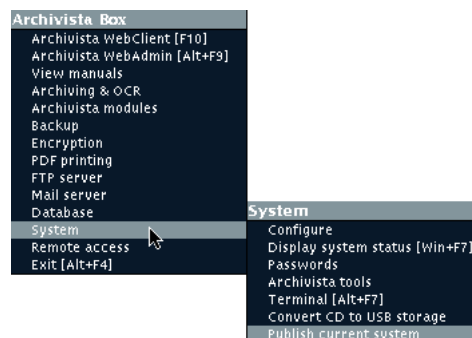
In the background of ArchivistaBox MySQL databases are at work. In order that all changes at database level can be traced so-called binary log files are created. In general we must reckon that for each piece of information added we need almost the same amount of space temporarily for the binary log files. As ArchivistaBoxes save mostly image data, a similarly large mass of data may be created in the binary log files. For this reason you can delete these log files with the help of this menu item at regular intervals (over months/years).

25.13.6 Clear slave binlog

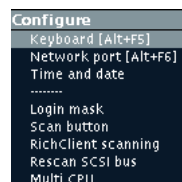
This option is needed only if you work with a master/slave configuration. Underlying this option is the same principle as in the case of 'Clear master binlog'.

➡ After the binary log files have been deleted a check if the master/slave configuration is still intact is mandatory. In case of doubt the master/slave configuration is to be set up afresh.

25.14 System



25.14.1 Configure



25.14.1.1 Keyboard (Alt+F5)

Choose the keyboard layout here. The following layouts are preset.

- US
- French
- German
- Italian
- SwissGerman
- SwissFrench
- SwissItalian

To choose a different keyboard layout click 'Other' and enter the abbreviation for your keyboard. When in doubt choose the US layout as all keyboards accept it.

25.14.1.2 Language

Here you can define the language you want to use.

25.14.1.3 Network port (F6)

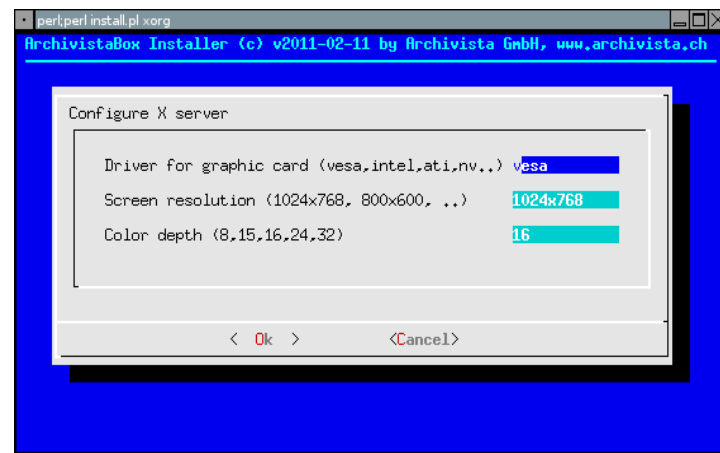
Here you define the network port. Please fill in IP address (CIDR notation), gateway and dns server.

➡ The CIDR notation is the short version for the submask. Example: Network with range from 192.168.0.0 to 192.168.0.255, the CIDR notation is /24. If you want to set the IP address to 192.168.0.100, just fill in 192.168.0.100/24.

25.14.1.4 Time and date

Here you determine system date and time.

25.14.1.5 Configure X server



When you start up ArchivistaBox, you will always get the 1024x768 screen resolution. This ensures that all current graphics cards and monitors may be used.

With this option you can enable the advanced settings of the computer's graphics card. Choose card type (e.g. vesa, ati, nv, Intel), resolution (1920x1080, 1440x900, 1240x1024) and colour depth (24-, 16-, 15-bit).

When you have confirmed the changes, you must restart the computer for the changes to take effect.

➡ If you have entered incorrect settings, the X server will be unable to start. In these cases you will be brought to the console. Log in as root user and run the following two commands:

```
cd /etc/perl
perl install.pl menu,de
```

You now can define new (correct) values.

25.14.1.6 Extended parameters

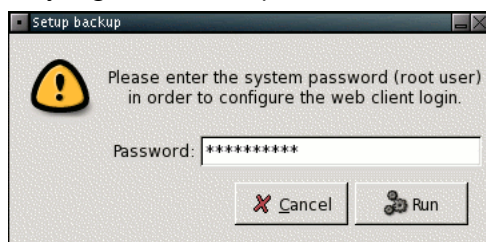


You can set the most important parameters using the advanced parameters. These include, in particular, network card and X server.

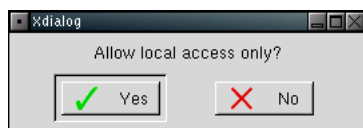
Next you can start the box again and a file manager (with root privileges) will be available.

25.14.1.7 Login mask

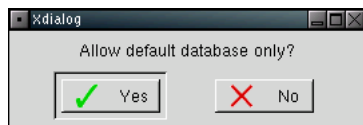
With this menu you can tailor the login mask of the WebClient – but not that of the WebAdmin tool – to your users' needs. Advantage: Users need not enter the name of the host each time they log in, for example.



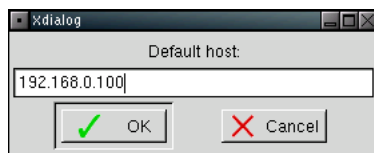
As always you will be asked for the password.



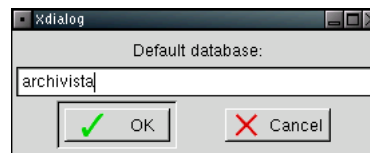
If access should be granted only from the local computer, click 'Yes'.



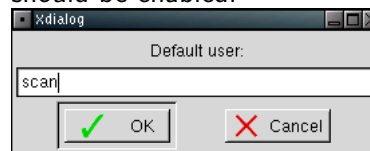
If always the same database is to be accessed, click 'Yes'.



You will be asked for the host on which the database is located.



You will be asked for the name of the database to which access should be enabled.



Here enter the user whose name should appear in the login mask.



In our example the fields for host and database are no longer displayed.

25.14.1.8 Scan button

This function enables you to use a numerical keypad to start the scanning process. It makes scanning a bit easier and more convenient.

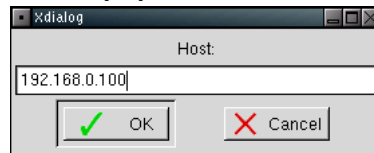


➡ The following procedure is no longer necessary for ArchivistaBox systems dating from 2013 and later because there the scan button is already active. Likewise, the scan button is active on ArchivistaBox Dolder if for scanning one of the keypad options

was purchased. Dolder systems that were purchased without the scanning module do not allow the capturing of documents with ArchivistaBox plus keypad.



As always you will be asked for the ArchivistaBox root password.



Then, enter the host on which the database lies to which you want to scan.

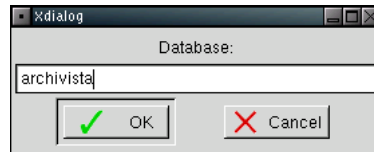
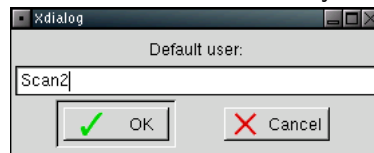
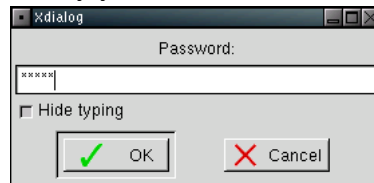


Illustration above: Here you must specify the database name.



Finally you enter the user...



...and the user password.

➡ Do consult 32.3, if you want to know more about the settings needed for remote scanning with an ArchivistaBox.

After these entries have been made, there are no further obstacles to your scanning with a numerical keypad.

➡ If the AccessLog function is turned on and you want to use the scan button, then you can only work with it if a connection to the WebClient has been established.

25.14.1.9 Admin-Benutzer (RichClient)

This function is no longer needed.

25.14.1.10 Configuring WebConfig

You can use this menu item to activate or deactivate the Web-Config application. Although WebConfig is activated as standard, you can deactivate it here or activate it again at a later stage.

25.14.1.11 Extended form for login

When you start the WebClient, WebAdmin or WebConfig, you will see on the upper left of the screen that there are direct links to the applications and handbooks. If you do not want these to be displayed, you can deactivate the 'Extended form for login'.

25.14.1.12 Determining the table structure for images

The ArchivistaBox usually saves the image data (which makes up approximately 90 to 95% of data) in a table. In the case of very large archives (between 50 and 100 Gbytes), it is recommended that you use this menu item to divide the tables that contain the image data into smaller units.

Use this menu item to login. The system then asks you how many folders you want to save in a table. Under normal circumstances,

you can enter the value '10' here. In other words, a new table is opened automatically for all ten folders.

Of course, you can enter any value between 1 and 100. If the level of colour is high, you should keep the number of files as small as possible (e.g. 5). If you only work with black and white images, the value can be higher (e.g. 20).

➡ You can use '0' to merge tables that you have already divided into one table. This option is available for reasons of compatibility. Remember to save your data before you divide the tables. If a power cut occurs during the conversion, the tables may be lost (although this is unlikely).

25.14.2 Display system status (Win+F7)

Here you can gain information about the system. You see, for example, which network port was allocated. This is important when the ArchivistaBox is used in a network.

25.14.3 Passwords

The ArchivistaBox has administrator passwords on three levels:

- Admin
- User 'archivista'
- User 'root'

The user named 'Admin' is designed for administration tasks in the WebAdmin tool, i.e. the creation of fields and input masks or the setting up of new users. User 'Admin' is an ordinary user with superuser (SYSOP-) rights **on the level of the database**.

User 'archivista' is designed for the 'external relations' of the ArchivistaBox. 'archivista' allows the exportation or importation

of data, for instance. Internally, in the database, it corresponds to user 'SYSOP'. 'SYSOP' cannot be edited in WebAdmin.

User 'root' possesses the most far-reaching rights both within MySQL as also for the entire system. Only user 'root' may create or delete databases or make system configuration changes.

25.14.3.1 Set system (root) password

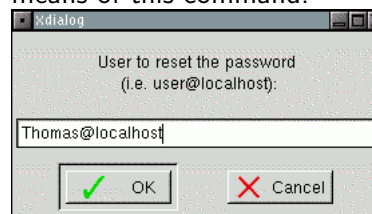
Run this command and set the new system (= 'root') password. User 'root' has rights both on the level of the operating system as on the level of MySQL. The password is the same and is defined here.

25.14.3.2 Set Archivista passwords

Run this command and set the new 'archivista' password. User 'archivista' possesses the same password as user 'SYSOP'.

25.14.3.3 Reset user password

Should a user have forgotten his/her password it can be reset by means of this command.



25.14.4 Archivista tools

25.14.4.1 Unlock documents

Whenever a document is being processed (e.g. when pages are added), it is locked for a very short time frame. This is to prevent two users from adding pages to the same document, for example.

Unfortunately, it may happen that a document was not unlocked automatically because the job was never finished (e.g. due to power failure or to the scanner being turned off during scanning). In this case, the document remains locked.

This auxiliary program lets you unlock the document in question manually in order that you may continue working with it.

25.14.4.2 Log-Tabelle löschen

25.14.4.3 Clear log table

By means of this submenu item you can delete the OCR log file. (For information on how to look at the log file in the first place compare section 25.6.5.2.) Removing the log items can be helpful if the text recognition module does not seem to work although it is turned on. In certain cases a 'warped' log entry may obstruct the correct functioning of the OCR software.

25.14.4.4 Remove documents or folders

With the help of this menu item you can delete certain documents or folders for good even if the documents have already been archived. Please use this command with the necessary prudence. In principle you need this menu item to delete documents definitively only after the retention period according to law has elapsed.

25.14.4.5 Restart OCR

If for some reason OCR is not running, use this function to restart it.

25.14.4.6 Register OCR

This function is no longer needed. Text recognition is activated automatically since 2012.

25.14.5 Terminal (Alt+F7)

This menu item is reserved exclusively for the developers of ArchivistaBox.

25.14.6 Online-Update

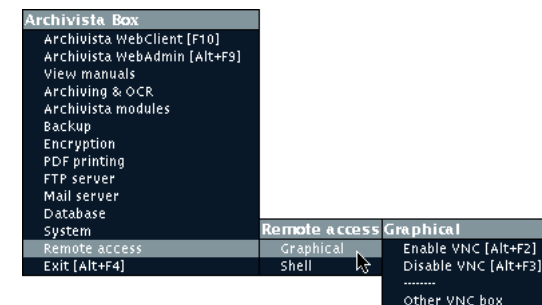
You can use this menu item to update ArchivistaBox to the latest version directly via the Internet. After clicking this menu item, you have to confirm the process. At the end, a message will ask you to restart ArchivistaBox.

25.14.7 Publish current system (installed ArchivistaBox)

This menu item is used to copy an existing ArchivistaBox (incl. database) to a DVD or to an external hard disk, respectively. Notwithstanding, since 2012 archives can be published directly in the WebClient. For more information on how archives are published see section 28.

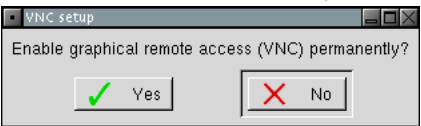
25.15 Remote Access

25.15.1 Graphical

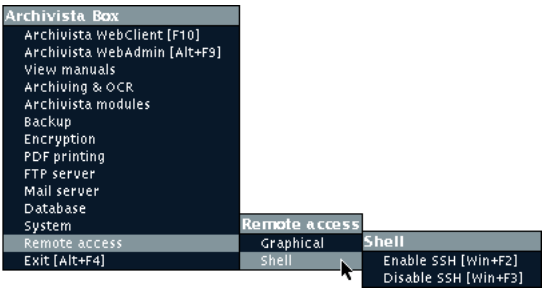


25.15.1.1 Enable VNC (Alt+F2)

With 'Enable VNC' you determine that the ArchivistaBox with which you are currently working can be accessed from other computers in the network by means of VNC. 'Disable VNC (F3)' will close the graphical remote access again. The command 'Other VNC box' serves to 'remotely enable remote access' so to speak. Remote access is a maintenance account and should not normally be open. However, it is possible to keep it open permanently in special cases. Click 'yes'. (Default value is 'No'.)

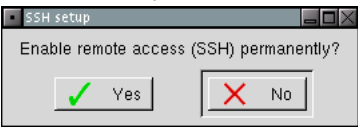


25.15.2 Shell



For a secure connection to the console you run 'Enable SSH (Win+F2)'. With 'Disable SSH (Win+F3)' you close the connection again.

Remote access is a maintenance account and should not normally be open. However, it is possible to keep it open permanently in special cases. Click on 'Yes' in the dialog displayed below. (Default value is 'No').



25.16 Exit (Alt+F4)

With this command you shutdown ArchivistaBox.

25.17 Function keys

The most important functions of the ArchivistaBox can be accessed by function keys:

Alt+F1	Documentation in English
Win+F1	Documentation in German
Alt+F2	Enable remote access (VNC)
Win+F2	Enable remote access (SSH)
Alt+F3	Disable remote access (VNC)
Win+F3	Disable remote access (SSH)
Alt+F4	Shutdown ArchivistaBox
Alt+F5	Configure keyboard
Alt+F6	Configure network port
Alt+F7	Open terminal window
Win+F7	Display system status
Alt+F8	Unlock documents
Alt+F9	Start WebAdmin
Alt+F10	Start WebClient
Alt+F11	Start WebConfig
F11	Turn on/off full screen mode
F12	Quit web browser (kill it)

Part VII

ArchivistaVM

26 Server virtualisation with KVM

26.1 Introduction

Historically, ArchivistaBox has been created as a document management box (DMS). The basic idea behind this is that a standard computer is cheaper than the work of a specialist, who would set up the system manually.

While a standard operating system can be set up in 15 to 30 minutes, the basic installation does not include specialised applications ready for operation. Server services, updates, client and backup programs have to be manually and individually set up.

With the growing complexity of these processes, it has become common practice to set up a server for each specialised application. This is where virtualisation comes into its own. Instead of setting up a physical server for each solution, several specialised applications are bundled in a type of cage, each with its own operating system, and operated virtually on a powerful server.

However, advanced virtualisation solutions are still considered neither simple nor inexpensive. In addition to established products, OpenSource products have been available for a number of years now, but these require extensive Linux knowledge and skills to configure and maintain. We created ArchivistaBox because we wanted to overcome these limitations. Consequently, we began the search for a virtualisation solution that would be easy to set up and thus present an alternative for our customers.

The first version of the ArchivistaBox for virtualisation was released by us in 2009. If initially the first solution was an updated Proxmox-Fork (see www.proxmox.org, today ArchivistaVM can no longer be compared with Proxmox. The focus is very different. Whilst Proxmox mutated to an all-encompassing data centre (with a current size of 550 MB), ArchivistaVM is designed to make vir-

tualisation a simple(r) technology. Not only has the ISO file been reduced considerably to 100 MB (which makes for extremely rapid starting), but ArchivistaVM also comes with completely automatic configuration - even when linking servers (clusters).

Under these conditions, ArchivistaVM performs 'class compliant'. It is based on Debian. And those who want to can install additional packages at any time using 'apt-get update' and 'apt-get install xyz'. One thing is worth noting: the ArchivistaBox (with the exception of virtualised guests) runs exclusively in the main memory (RAM). Those expecting an installation with user interaction will be just as disappointed as those expecting (as with all other solutions) to have to reboot following installation. Once the ISO / stick has been booted, that's it! When starting, the server is automatically configured, so that after approx. 20 or 30 seconds (1 - 2 minutes in the case of a cluster), the infrastructure is ready to use.

➡: The following chapter explains how to use ArchivistaVM, irrespective of whether you work with ArchivistaVM as a module of the ArchivistaBox DMS solution or whether you use the ArchivistaVM module alone to run the server virtualisation.

26.2 System requirements

The ArchivistaVM module requires 64-bit AMD or Intel processors that have virtualisation chipsets. In principle, this is always the case with AMD processors. However, some Intel processors, in particular so-called entry-level CPUs (e.g. Atom), do not support virtualisation.

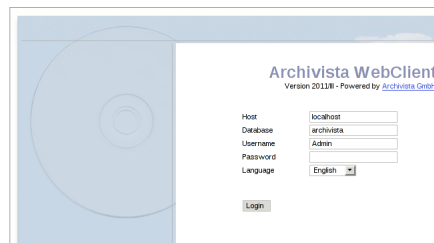
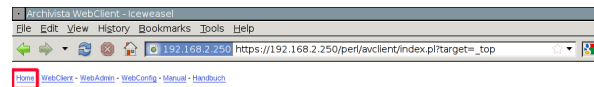
➡: Even if the processor is suited for virtualisation, it is possible that this has been disabled in BIOS. If this is the case, the appropriate options must be enabled in BIOS.

The recommended hardware configuration for virtualisation includes at least four core processors (CPUs), with high hard disk throughput and plenty of memory (RAM), as the resources must be divided by the number of running instances. We highly recommend hard drive arrays (hardware RAID and software RAID).

26.3 Login to ArchivistaVM

26.3.1 ArchivistaBox (WebClient) login

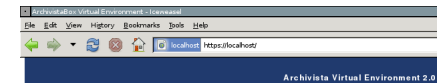
If you have installed ArchivistaBox, the login screen of the WebClient will appear during startup. Clicking on 'Home' will take you directly to the login screen of ArchivistaVM. For details see section 26.3.2.



➡: You can go directly to the ArchivistaVM module by entering the IP address of the ArchivistaBox (192.168.2.250, in the following example) and going to 192.168.2.250/index.htm. This will take you directly to ArchivistaVM.

26.3.2 Login to ArchivistaVM

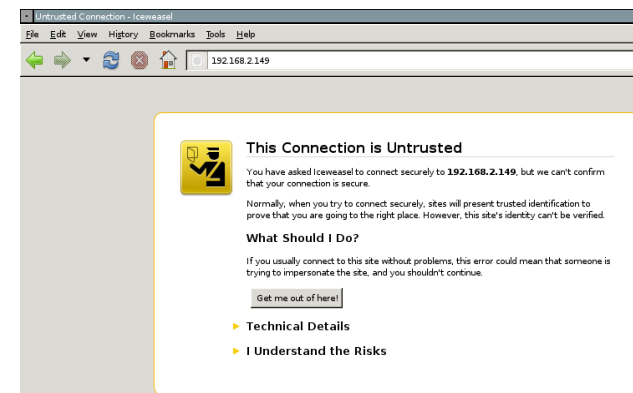
If you have installed the ArchivistaVM CD, you will be taken directly to the login mask of ArchivistaVM. Enter 'root' as user name and enter the root password.



➡: The default password is 'archivista'.

26.3.3 External login

If ArchivistaBox is connected to a network, you will be taken to an input screen when you enter the IP address of the ArchivistaBox together with '/index.htm'. For example, if the IP address is 192.168.2.250 you would enter 192.168.2.250/index.htm.



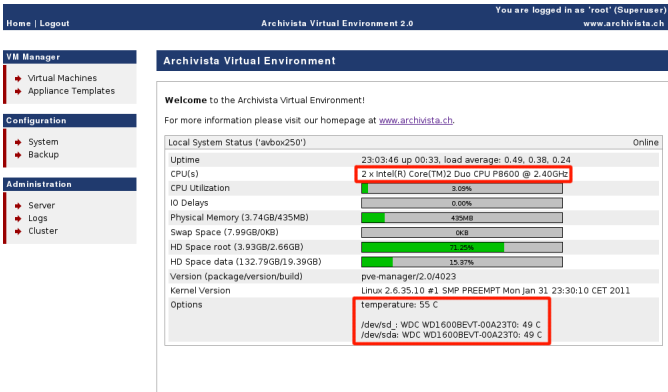
When you login for the first time, you will see a query regarding the certificate of the site. Add an exception to your browser so you can access the web interface of ArchivistaVM.

➡: Clicking on 'Home' on the top right of the login screen will take you to the WebClient (DMS-part). The same applies in reverse when you see the login screen of the WebClient. Clicking on the top left button 'Home' will take you to ArchivistaVM.



26.4 System overview

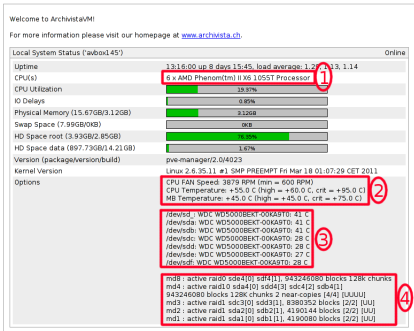
After you log in, you will see the system overview, which provides important information on the ArchivistaBox status.



The most important values are the number of cores and the system temperature of the processors and hard drives.

26.4.1 Extended status report (software raid)

The information you will receive depends on the type of ArchivistaBox.



In the above example, under point 1 you will see the number of CPUs (6 cores). Under point 2, you will see system temperature data. Finally, point 3 shows the temperature range of the hard drives.

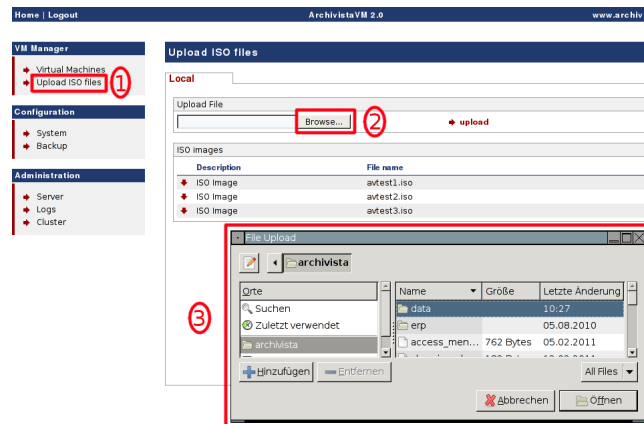
You have read it correctly - this system has a total of 6 hard disks configured in a RAID array (software RAID). If more than one hard drive is shown under point 3, then you will find an overview of the present status of the RAID array under point 4.

26.5 VM Manager

Before you can run the operating system in a virtual mode, you will need the installation CD and/or DVD of the guest system. You will use the physical CD/DVD to create a so-called ISO file.

26.5.1 Uploading ISO files

You can upload the ISO file via the web interface. For details please see the following screenshot.



Select 'Upload ISO files' (point 1), then click on 'Browse' (point 2; this button could also be called 'Upload'). Finally, select the appropriate ISO file (point 3) to upload it to ArchivistaVM.

26.5.1.1 ISO files over 2 GB

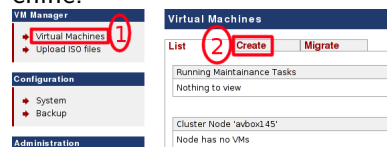
At present, most Web browsers cannot upload files larger than 2 Gb. However, ISO files often exceed 2 GB. In order to upload these to the ArchivistaBox, it is best to use a SCP copying program, which is WinSCP for Windows.

The ISO files must be uploaded to the following folder:

```
/var/lib/vz/template/iso
```

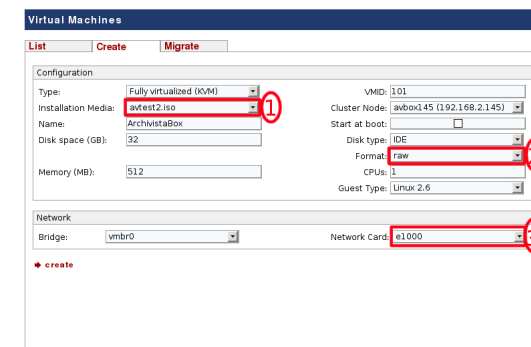
26.5.2 Virtual machines

Now you can create a virtual instance. Select 'Virtual Machines' and then click on 'Create' to set the basic parameters of the machine.



Now you can set the basic parameters of the machine. Please take into account the following points:

- ISO file: Under point 1, set the installation disc (ISO file).
- Hard drive: In point 2, you can determine the format of your hard disk. Choose between 'qcow2' and 'raw'. If you select 'qcow2', all changes to the virtual disk will be logged. This option is particularly useful if you want to test systems. This allows you to save a copy of the current state, which you can go back to later if necessary (e.g. if an update fails). This is not possible when you select the raw format. However, the advantage of the raw format over qcow2 is its higher efficiency under a higher load and over a longer period of time.
- Network card: Under point 3, set the virtual card that can be addressed later by the operating system. You have the choice between the rtl8139 or e1000 network card. While the first option does not require any drivers even in older operating systems, the e1000 is faster (1Gb card), but usually requires specific drivers.



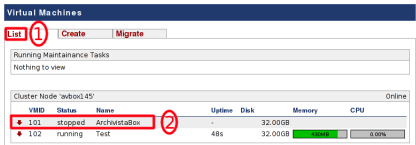
⚠ Important: To save your settings, you have to click on the red 'create' button at the bottom of the screen. You will receive a confirmation that the instance has been set up.

If you look at the text of the message, you will see that the command corresponds to what is necessary to create an appropriate instance on the console.



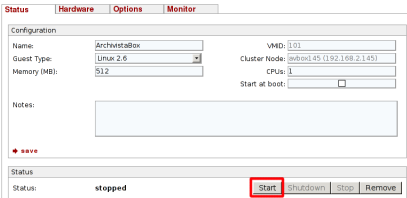
26.5.2.1 Start installation

If you click on the list under point 1, you will see an overview of all instances. In general, you will find the instance which was set up most recently at the bottom of the list. In our case, it is the instance called 'ArchivistaBox' under point 2.

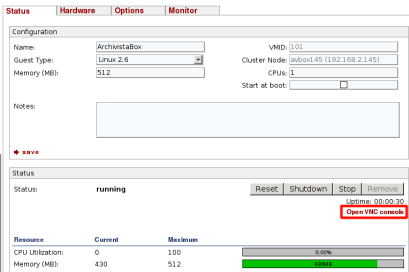


➡: The entries which have green bars next to them are instances which have been started.

In our case, we want to start the new instance, 'ArchivistaBox'. Therefore, we will click on the entry at the bottom of the list. The following form appears:



You can now click on 'Start' to start the instance. The form changes slightly:

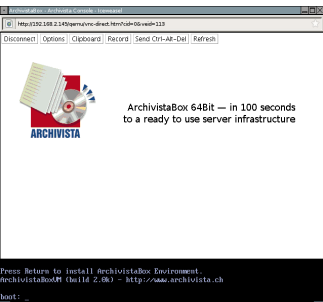


The button 'Start VNC console' appears. Now click on this button to start the virtual screen of the instance.



You may have to confirm the so-called digital signature first.

After you click on 'Run', you will see the installation screen of the instance.



You can now install the operating system (easily via the Web browser).

➡: Once you complete the operating system installation, you are likely to see a message asking you to restart the system. Please do not restart the computer at this stage, as it is sufficient to restart the virtualised instance. In this case, consider the virtualisation server (ArchivistaVM) quite simply to be the server room, and the

virtualised instance the server. Taking this into account, it would not make sense to pull the plug for the server room in order to restart a computer.

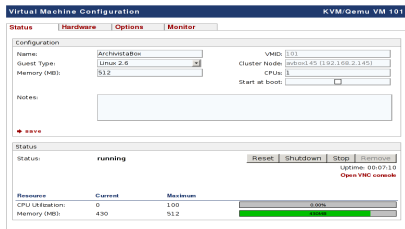
26.5.3 Configure instances

Whenever you click inside the VM Manager on 'Virtual machines', you will see the list of all instances (machines).

In order to work with an instance, click on the desired entry in the list (exactly as you did during the installation as described in section 26.5.2.1).

For example, you can increase or decrease the memory allocation (RAM), add new hard drives, integrate CDs and much more. In the following sections, we will discuss 'Status', 'Hardware', 'Options' and 'Monitor'.

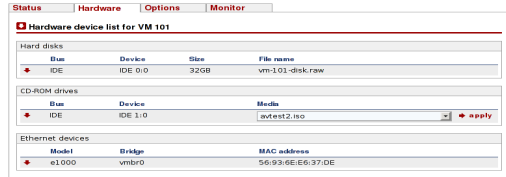
26.5.3.1 Status



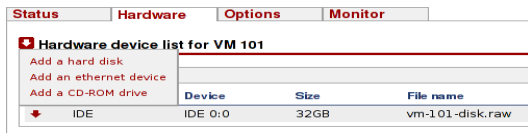
You can manage an instance under 'Status'. This includes session 'Start', 'Reset' and 'Exit'. The detailed status information about a running instance can also be found here.

Whenever you click on the 'Start' button, you will see the option 'Start VNC Console'. This will open the virtual screen of a running instance.

26.5.3.2 Hardware



Under hardware, you can specify hard drives, CD drives and network cards and later modify them. To add a device, click on the arrow icon to the left of the 'Hardware list ...'



26.5.3.3 Add hard disk

First, we will set up a new hard drive. You will see the following screen:



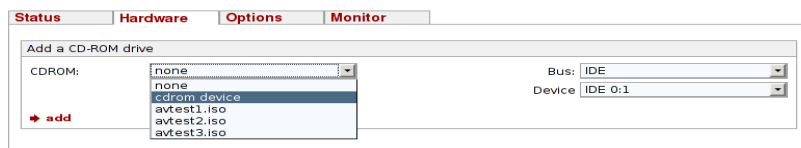
In addition to the size of the disk, determine its type (Bus). Typically, this will be IDE. In the case of virtualised hard drives, SCSI does not provide a speed advantage. Use VIRTIO for the highest throughput, but you can only use these hard drives in the virtualised instance if you have the appropriate VIRTIO driver.

Select 'qcow2' or 'raw' in the format field. In the case of 'qcow2', the virtualised hard disks increase with every change. In principle, they are versioned. This allows you to create a backup of the current state, which you can go back to later if necessary. Use the 'raw' format if you do not want to create versioned disks. We recommend the 'raw' format for high throughput.

You have the option to enter the device number in the 'Device' field, but you can normally leave ArchivistaVM to enter this information automatically.

26.5.3.4 Add CD-ROM drive

This allows you to select CD ISO files, and - if a physical drive is available - put this drive to use.



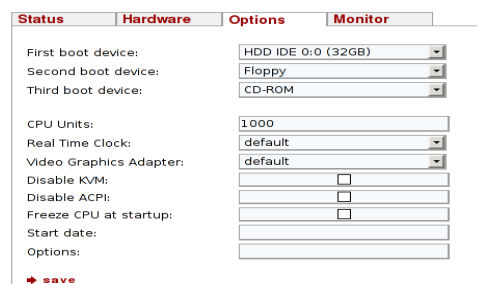
Click on 'Create' to add the drive to the instance.

26.5.3.5 Add a network card



Under 'Add Ethernet Device', a new network card can be added. Of course, you can assign several devices to one instance.

26.5.3.6 Options



- Boot sequence: Define the sequence of drives when you start the instance.

- CPU units: With these options, you can assign a higher or lower priority to an instance. Usually, the default values can be left unchanged.
- Real time clock: Choose between Universal Time (UTC) and local time.
- Graphics card: You can choose between a VGA and Cirrus card. The Cirrus card offers higher resolutions.
- Disable KVM: If you enable this option, the instance with pure software emulation is started. This may be necessary or useful for older instances.
- Disable ACPI: It may be necessary for you to turn off ACPI for older operating systems.
- Stop CPU after start: You can start an instance, but stop the CPU afterwards. This is useful if you would like to work with an instance at a later stage.
- Start date: You can specify to start an instance at a particular time.
- Options: Not all settings can be specified through the web interface. You can use this field to configure options which cannot be set through the web interface. Please note that an incorrect setting may lead to an instance not working or no longer working properly.

26.5.3.7 Monitor

```

Status Hardware Options Monitor
drive_add pci_addr=[<domain>]:<bus>:<slot>
[if=file,if-type=<bus>,bus=<n>
l,unit=ml,media=dl,index=<sl>
l,cyls=c,heads=h,secs=s,l,trans=t]]
[snapshot=on|off],cache=on|off] -- add drive to PCI storage controller
pci_add pci_addr=auto[<domain>]:<bus>:<slot> nic:storage:host [vlan=<n>],macaddr=addr[,mo
del=type]] [if=file,if-type=<bus>,bus=<n>... thost=02:00:0f,name=string[,dma=none] -- hot-ad
d PCI device
pci_del pci_addr=[<domain>]:<bus>:<slot> -- hot remove PCI device
host_net_add tap:user[:socket]:deidump [options] -- add host LAN client
host_net_remove vlan_id name -- remove host LAN client
host_net_redir [tcp|udp]:host-port[:guest-host]:guest-port -- redirect TCP or UDP connections
[from host to guest (requires -net user)]
balloon target -- request VM to change it's memory allocation (in MB)
set_link name up|down -- change the link status of a network adapter
watchdog_action [reset|shutdown|poweroff|pause|debug|none] -- change watchdog action
acl <command> <aclname> [<match>] [<index>]]
-- acl show vnc.username
acl policy vnc.username deny
acl allow vnc.username fred
acl deny vnc.username bob
acl reset vnc.username
cpu_set cpu [online|offline] -- change cpu state
qm>

```

ArchivistaVM or KVM has a monitor mode, which can be used to monitor a running instance or to make modifications. It also allows you to send keyboard entries to an instance.

🔗: The monitor mode is very powerful. The options are described also in the KVM handbook and other instruction manuals. If you have difficulties, your first step should be to type `help`.

26.5.3.8 Backup and load instances

If you configure hard disks in the qcow format, you will be able to create backups of running sessions on an ongoing basis. To create a backup, use the command `savevm name`. With `loadvm name`, you can activate a backed up state, and you can use `info snapshots` to obtain an overview of the backups.

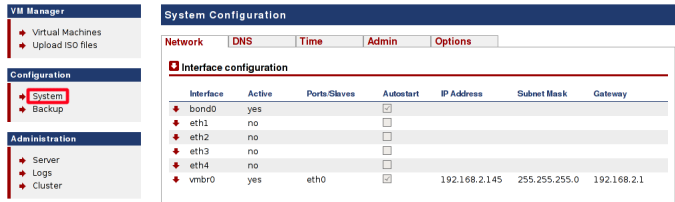
```

Entering Qemu Monitor for VM 102 - type 'help' for help
qm> savevm start
qm> loadvm start
qm> info snapshots
Snapshot devices: ide0-hd0
Snapshot list (from ide0-hd0):
ID      TAG      VM SIZE      DATE      VM CLOCK
1       start    138M 2011-04-02 14:18:15  00:29:34.289
qm>

```

26.6 Configuration

26.6.1 System



Under this menu item, you can configure ArchivistaVM.

🔗: You can set many of the options in several places on our ArchivistaBox. The settings for the network card, for example, can be configured through WebConfig or the system menu.

26.6.2 Backup

26.6.2.1 There is no secure data backup concept currently available for virtualisation!

In virtualisation, so-called 'hot-swap backups' (other names such as 'snapshots' are also commonly used) are almost always implemented. 'Hot-swap' means that, prior to backing up, the machine is put into the static condition, so that backing up (without guest shutdown) can then be carried out.

In order that guests can be retrieved 'without problems' at some later point, the main memory (RAM) and the hard drive files are normally backed up in such a way that the guests are temporarily 'frozen in'. An image of the main memory (RAM) is then created and backed up together with the contents of the hard drive file(s). During restore, the backed-up content of the main memory (RAM) and the hard drive files are shifted back to the time of the original backup, i.e. the guests are back-dated to the point in the operation when the hot-swap data backup was implemented.

This ensures that the exact conditions that existed at the time of the data backup are recreated. However, such backups cannot be regarded as static, because during hot-swap operations, no checks are carried out as to whether the data actually exists in a static condition. In the worst case scenario, the backup is in a state where the operating system itself has stalled. Alternatively, the backed-up files in the database server are corrupted; in the best case scenario, these files can be retrieved, though retrieval is usually extremely time-consuming.

For database applications (which are the rule in the server environment), plug-ins are often installed in the guests, so that the databases are briefly paused prior to the creation of the memory image (for example); this, however, requires plug-ins to be available and also that the corresponding versions, the virtualisation software and the guest software are compatible. So, what can go wrong? Very simply, the virtualised guests can only be backed up if the virtualisation software provider makes the operating system plug-ins available. Or to put it another way, if there is no plug-in being used for the operating system, no accurate data backups can be carried out. Ultimately, this is sold as 'certified for XYZ', but it should be labelled 'limited to XYZ'.

26.6.2.2 The ArchivistaVM concept

Things can be a lot simpler than the process described above. The host (virtualisation software) and the guest (virtualised operating system) operate autonomously, the guest is briefly shut down during data backup, data backup is executed and the guest is rebooted. If the guests are redundant on the hard drive, the time during which the guest is unavailable can be extremely short. Guest shutdown, 2nd hard drive unmounted, guest booted, backup executed, 2nd hard drive re-mounted.

If no guest shutdown is carried out (in the Linux environment this is generally frowned upon, long uptimes are the pride of any administrator), the relevant scripts in ArchivistaVM can be started outside the guest before and after data backup, which thereby ensures that the data in the guest is in a static condition. But if we're being honest, even Linux operating systems are nowadays much more readily updated than they were previously, and here too it is becoming increasingly desirable to backup the entire guest, so as not to be in the position of having the wrong version of the database server when retrieving on some future 'Day X'.

The concept used by ArchivistaVM is sadly not used for virtualisation solutions, although instances are 100% static and autonomous only in this concept. The concept needs no additional data backup software, as many instances as required can be backed up retrospectively and there is no guest-dependency - in a nutshell, these are standard compliant hard drive files which, if necessary, can be booted or opened offline.

➡ Note: Just to make sure that we all completely understand each other, ArchivistaVM can, of course, implement trouble-free hot-swap backups in the qcow2 format, but these backups contain precisely the same defects as the other solutions, so we won't highlight these further!

26.6.2.3 Backing up guests

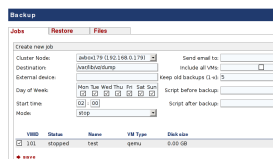
Of course, guests (instances) can only be backed up if they have already been set up; if nothing has been set up, then there is nothing to back up. If this is the case, the data can be backed up in ArchivistaVM to the left of the 'backup' menu item and in 'jobs' by going down and clicking on the 'create new job' arrow.



The data backup options can now be established. Note the 'Destination' and 'External device' fields. Provided no external device (hard drive) is present (which is the case in the following first attempts), the data is backed up into this directory on the internal hard drive.

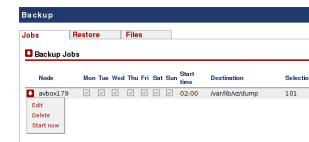
➡ Note: If the data needs to be saved to an external hard drive at some later point, then naturally, in the context of these instructions, it would be advisable to first try out the simpler variant (no external drive necessary).

The 'Keep old backups (1-x)' field deserves special mention at this point. This allows specification of the number of backups before the last (oldest) backup is overwritten.



➡ Note: In order to back up the data onto an external hard drive, the drive identifier must first be entered (e.g. /dev/sdb1, /dev/sdc1) under 'external device'. In this case, the target directory merely serves as the mounting point for the external device; the guests themselves are not saved in this target directory, but directly onto the external device. The external drive identifier can be called using the 'restore' tab.

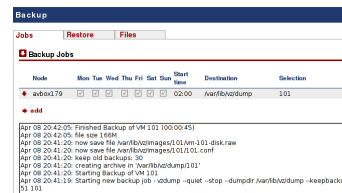
➡ Note II: In contrast to earlier versions of the ArchivistaVM, NTFS-formatted drives can be used with the current version. However, it should be noted that when using NTFS-formatted drives, no restore-on-the-fly is possible.



Data backup can now be set to the desired time (in this case 02:00 am). The job should, of course, be tested immediately. This is done by going down, clicking on the arrow for the corresponding job and then selecting 'Start now'. This triggers a prompt, asking whether data backup should be started. Please confirm.



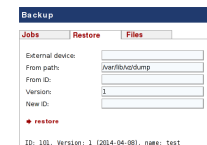
The data backup will now be executed. The log files can be monitored under the 'data backup' menu item.



Once the operation is complete, the 'finished backup' message is displayed.

26.6.2.4 Retrieving backed-up data

In order to retrieve previously backed-up data, use the 'restore' tab next to 'jobs' in the 'data backup' sub-menu. This allows viewing of the existing backups.



In order to retrieve a backup, its ID, the correct version number and a new ID for the restored data must be entered. In our example, ID 101 is to be retrieved using Version 1 and restored to ID 201.

Backup

JobsRestoreFiles

External device:

From path:

/var/lib/vz/dump

From ID:

101

Version:

1

New ID:

201

restore

ID: 101, Version: 1 (2014-04-08), name: test

The operation can now be executed using the 'Restore' tab. Here too, status reports are available under the 'jobs' tab. Once the 'restore successful' message is displayed, the operation is complete.

Backup

JobsRestoreFiles

Backup Jobs

Node	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Start time	Destination	Selection
avbox179								02:00	/var/lib/vz/dump	101

add

Apr 08 20:50:07: restore successful
Apr 08 20:50:07: restore configuration to /etc/qemu-server/201.conf
Apr 08 20:49:28: space needed: 179436, available: 31667232
Apr 08 20:49:28: restore qemu-server image 101 using ID 201
Apr 08 20:42:05: Finished Backup of VM 101 (00:00:45)
Apr 08 20:42:05: file size 160M
Apr 08 20:41:20: now save file /var/lib/vz/images/101/vm-101-disk.raw
Apr 08 20:41:20: now save file /var/lib/vz/images/101/101.conf
Apr 08 20:41:20: keep old backups: 30
Apr 08 20:41:20: creating archive in /var/lib/vz/dump/101
Apr 08 20:41:20: Starting Backup of VM 101
Apr 08 20:41:19: Starting new backup job - v2dump --quiet --stop --dumpdir /var/lib/vz/dump --keepbackup 30 101

➡ Note: For large guests (instances), this operation can take several hours. Depending on the hard drive and interface used (USB2/USB3), average retrieval speeds of 50 - 200 MB per second can be expected.

➡ Note II: Instance retrieval cannot be started unaltered whilst another instance is in progress, since this would mean the same Mac- and IP addresses being used twice. The Mac-address can be altered using the 'hardware' tab, by removing the network card and then replacing it with a new Mac-address. The guest can then be started and the IP-address(es) changed. Only then can the two instances be started (if, of course, this is actually necessary).

26.6.2.5 Backing up onto external devices (hard drives)

In order to backup onto an external hard drive, the correct identifier (e.g.: /dev/sdb1, /dev/sdc1) for the job definition must be entered. The simplest way is to call up the identifier using the 'restore' tab in the connected (not the mounted) drive. Provided the drive is connected, the identifier will be visible under 'External device' (in our example: /dev/sdb1).

Backup

JobsRestoreFiles

External device:

/dev/sdb1

From path:

/var/lib/vz/dump

From ID:

Version:

1

New ID:

restore

A new backup can now be created under 'Jobs', by entering the identifier under 'External device'.

Backup

JobsRestoreFiles

Create new job

Cluster Node:

avbox179 (192.168.0.179)

Destination:

/var/lib/vz/dump

External device:

/dev/sdb1

Day of Week:

Mon Tue Wed Thu Fri Sat Sun

Start time:

02 : 00

Mode:

stop

Send email to:

Include all VMs:

Keep old backups (1-x):

2

Script before backup:

Script after backup:

VMID	Status	Name	VM Type	Disk size
101	stopped	test	qemu	0.00 GB
201	stopped	test	qemu	0.00 GB

save

➡ Note: Provided that an entry exists under 'external device', the backup will immediately be executed onto an external hard drive. If the entry is inconsistent, the backup cannot be executed; this is then reported in the log file.

➡ Note II: The two options 'stop' and 'cluster' are available under 'mode'. The 'cluster' option allows backups to be created in clustered servers (a minimum of two ArchivistaVM servers). An instance to be backed up on the first node ends up on the last node, all remaining instances end up on the previous node. Here it must be ensured that the hard drive is connected to the correct node. For example: If an instance to be backed up onto the first

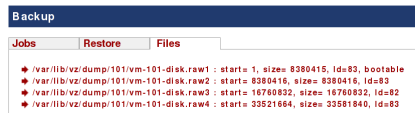
node of a double cluster ends up on the hard drive of the second node; this suggests that the hard drive connection to the first node is non-functional.

26.6.2.6 Restoring files

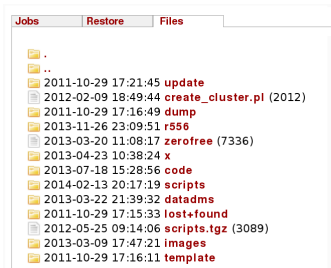
Click on the 'files' tab under 'data backup' in ArchivistaVM. This should list all the available backed-up RAW files.



Clicking on a file should now allow the available partitions to be viewed.



Any partition can now be opened directly. An example is given below.



Entries that have a number in brackets following the file name [e.g. zerofree (7336)] are files. All the remaining entries are folders. Clicking on a folder allows the associated sub-folders to be opened. Clicking on a file prepares the file for download.

➡ Note: Clicking on the 'Files' tab immediately displays the top level (i.e. all the RAW files).

26.7 Administration

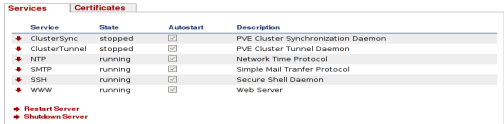
You can manage individual ArchivistaVM computers under 'Administration'.

26.7.1 Server

Under 'Server' you will find 'Services' and 'Certificates'.

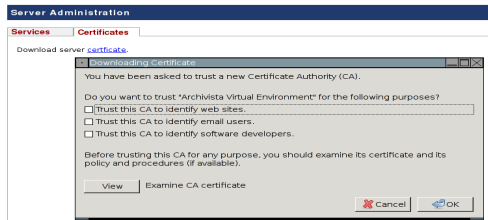
26.7.1.1 Services

This allows you to turn the important ArchivistaVM services on and off. You can also shut down and restart the ArchivistaVM computer.



26.7.1.2 Certificates

When you first access the ArchivistaBox from a remote computer you will be asked to confirm the certificate. Using this menu option, certificates can be obtained and saved to the computer or web browser. After that you will no longer be asked to confirm the certificate.



26.7.2 Logs

26.7.2.1 Tasks

This shows an overview of recently executed commands.

Tasks Syslog					
Running Maintenance Tasks					
Nothing to view					
Finished Maintenance Tasks					
Command	Start time	User	CID	VMD	
start	Sat Apr 2 13:56:56 2011	root	0	101	
start	Sat Apr 2 13:48:41 2011	root	0	102	
destroy	Sat Apr 2 13:32:39 2011	root	0	124	
destroy	Sat Mar 26 20:11:44 2011	root	0	103	
destroy	Sat Mar 26 20:11:40 2011	root	0	104	
destroy	Sat Mar 26 20:11:36 2011	root	0	105	
destroy	Sat Mar 26 20:11:31 2011	root	0	106	
destroy	Sat Mar 26 20:11:27 2011	root	0	107	
destroy	Sat Mar 26 20:11:23 2011	root	0	108	

26.7.2.2 Syslog

Under this menu, you can view some important log files.

Tasks Syslog	
Realtime Logfile Query (last 100 lines)	
Filter expression	Service: Apache
<pre> 127.0.0.1 - - [02/Apr/2011:14:32:44 +0200] "GET /images/iarrrdown.png HTTP/1.1" 200 419 "http://localhost/" 127.0.0.1 - - [02/Apr/2011:14:32:44 +0200] "GET /images/iarrrdown.png HTTP/1.1" 200 408 "http://localhost/" 127.0.0.1 - - [02/Apr/2011:14:32:44 +0200] "GET /images/iarrrright.png HTTP/1.1" 200 408 "http://localhost/" 127.0.0.1 - - [02/Apr/2011:14:34:08 +0200] "GET /javascript/prototype/prototype.js HTTP/1.1" 200 32016 127.0.0.1 - root [02/Apr/2011:14:34:08 +0200] "GET /server/certs.ite HTTP/1.1" 200 1546 "http://localhost/" 127.0.0.1 - - [02/Apr/2011:14:34:08 +0200] "GET /css/pve.css HTTP/1.1" 200 1303 "http://localhost/server/" 127.0.0.1 - - [02/Apr/2011:14:34:08 +0200] "GET /css/pve.js HTTP/1.1" 200 3409 "http://localhost/server/" 127.0.0.1 - - [02/Apr/2011:14:34:08 +0200] "GET /images/hlink.gif HTTP/1.1" 200 322 "http://localhost/" 127.0.0.1 - - [02/Apr/2011:14:34:08 +0200] "GET /images/arr-right.png HTTP/1.1" 200 408 "http://localhost/" </pre>	

26.8 Cluster

A cluster is a group of several ArchivistaVM servers, which can communicate with one other. In the case of ArchivistaVM, at least one computer is a master, to which nodes can be added. The master and the nodes have to be configured on the console.

🔑: The easiest way to get to the console is by keying in 'Alt+F7' on the ArchivistaBox. You will have to enter the archivista password, and you will also need root privileges. By entering `su` and the appropriate root password, you will be taken to the root console (identified by the # sign). Alternatively, you can log in to the ArchivistaBox from another computer with SSH (remote console).

26.8.1 Configuring a cluster with pveca

Type `pveca` on the console. You will receive an overview of the management program to create or edit the clusters.

```

avbox250a:/home/archivista# pveca
ERROR: no action specified
USAGE: pveca -l          # show cluster status
       pveca -c          # create new cluster with localhost as master
       pveca -s [-h IP]  # sync cluster configuration from master (or IP)
       pveca -d ID       # delete a node
       pveca -a [-h IP]  # add new node to cluster
       pveca -m          # force local node to become master
       pveca -i          # print node info (CID NAME IP ROLE)
avbox250a:/home/archivista#

```

26.8.2 Create the master

In order to set up a master computer for the cluster, enter `pveca -c` on the console. The current computer will become the master of the (new) cluster.

```

avbox250a:/home/archivista# pveca -c
cluster master successfully created
avbox250a:/home/archivista# pveca -l
CID---IPADDRESS---ROLE-STATE-----UPTIME---LOAD---MEM---ROOT---DATA
1 : 192.168.2.250   M    A              10:40   0.76   30%   90%   25%
avbox250a:/home/archivista#

```

Verify the IP address of the master by entering `pveca -l`.

26.8.3 Add nodes

Now you must set up the console on the ArchivistaVM computer that you want to add to the master as a node. In our example, we will add the node 192.168.2.252 to the master 192.168.2.250.

To do this, use the command `pveca -a -h IP`, which is in our case:

```
pveca -a -h 192.168.2.250
```

The following screen, or a similar one, should appear:

```

avbox252:~# pveca -a -h 192.168.2.250
Generating public/private rsa key pair.
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
3d:3e:e0:be:4c:b6:2a:6b:5a:fa:8f:6c:63:27:51:26 root@avbox252
The key's randomart image is:
+---[ RSA 1024]-----+
|
|      E o .
|      + S o
|      . . o .
|      . . + o
|      +B . = . .
|      o*=Bo.=.
+-----+
The authenticity of host '192.168.2.250 (192.168.2.250)' can't be established.
RSA key fingerprint is 61:46:d8:56:9d:ac:a7:43:5a:c0:ab:84:e2:75:3c:1c.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '192.168.2.250' (RSA) to the list of known hosts.
root@192.168.2.250's password:
cluster node successfully created
avbox252:~#

```

⚠: Please note that the node can be created only if you can login to the master computer with SSH (in our example 192.168.2.250). You can verify with `pveca -l` whether the node is running.

```

avbox252:~# pveca -l
CID---IPADDRESS---ROLE-STATE-----UPTIME---LOAD---MEM---ROOT---DATA
1 : 192.168.2.250 M A 10:49 2.09 30% 90% 26%
2 : 192.168.2.252 N S 09:48 1.87 81% 27% 15%
avbox252:~#

```

26.8.4 Cluster Management in Web interface



If you now log into ArchivistaVM again or click on the 'Home' button, you can directly access the cluster information via the web interface.

ArchivistaVM										
Welcome to ArchivistaVM!										
For more information please visit our homepage at www.archivista.ch										
Hostname	IP Address	Role	State	Uptime	Load	CPU	KDElay	Memory	Disk	
jetbox250	192.168.2.250	Master	active	1 day 00:32	0.24	11%	0%	26%	35%	
avbox252	192.168.2.252	Node	active	00:12	0.34	11%	0%	40%	19%	

Local System Status ('avbox250')										
Uptime 15:05:11 up 1 day 00:32, load average: 0.24, 0.70, 0.62										
CPU(s) 2 x Intel(R) Core(TM)2 Duo CPU P8600 @ 2.40GHz										
CPU Utilization 11.87%										
IO Delays 3.69%										
Physical Memory (3.74GB/1.35GB) 3.95GB										
Swap Space (7.99GB/0MB) 0B										
HD Space root (8.93GB/2.80GB) 0B										
HD Space data (132.79GB/44.19GB) 35.09%										
Version (package/version/build) pve-manager/2.0/4023										
Kernel Version Linux 2.6.35-11 #1 SMP PREEMPT Fri Mar 18 01:07:29 CET 2011										
Options										
/dev/md_0: WDC WD1600BEVT-00A23T0: 45 C										
/dev/sda: WDC WD1600BEVT-00A23T0: 45 C										

⚠: Make sure that all cluster machines are set to 'active' under 'Status'. If this is not the case, the cluster is not in an operationally ready state.

26.8.4.1 Open instance in clusters

If you open a cluster network on a new virtual machine, you will now see the new option 'Cluster node', where you can select the ArchivistaVM server on which you wish to run the machine.

List		Create	Migrate
Configuration:			
Type:	Fully virtualized (x86)	VMID:	109
Installation Media:	cdrom device	Cluster Node:	avbox252 (192.168.2.250)
Name:		Start at boot:	avbox252 (192.168.2.250)
Disk space (GB):	32	Disk type:	IDE
Memory (MB):	512	Format:	qcow2
		CPU(s):	1
		Guest Type:	Other

26.8.4.2 Overview of the instances

Now, when you click on 'Virtual Machine' (the overview), you will see an overview of all the virtualised instances throughout the cluster group.

Virtual Machines

List

Create

Migrate

Running Maintenance Tasks

Nothing to view

Cluster Node 'avbox250'

Online

VMID	Status	Name	Uptime	Disk	Memory	CPU
101	stopped	test	-	-	32.00GB	-
102	running	avtest2	18m	-	32.00GB	100%
103	stopped	test	-	-	32.00GB	-
104	stopped	avtest1	-	-	32.00GB	-
105	stopped	test1	-	-	0GB	-
106	stopped	avtest3	-	-	32.00GB	-
107	stopped	avvm	-	-	32.00GB	-
108	stopped	ArchivistaBox	-	-	32.00GB	-

Cluster Node 'avbox252'

Online

VMID	Status	Name	Uptime	Disk	Memory	CPU
109	running	avtest4	10s	-	32.00GB	80%

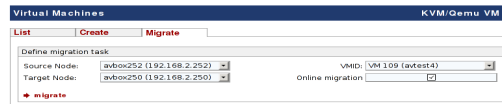
⚠: Once you are working with a cluster, you should always register on the master computer. No more new instances can be added to the node.

26.8.4.3 Migration of a machine

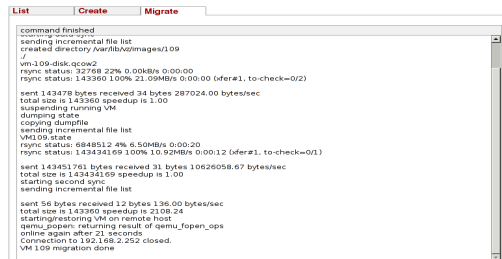
The advantage of cluster networks is that the corresponding instances may be directed (migrated) from computer X to computer Y in a restricted manner (even during operation). To do this, go

to 'Virtual machine' and then 'Migration'. You will now see the form to migrate an instance.

In 'Output node', select those ArchivistaVM servers from which you want to select a machine for the migration. Now you can select the desired instance with VMD. With 'Online migration' you can determine that the instance should be migrated while in operation. With 'destination node' you can determine to which computer the instance should be migrated.



Now click on 'migrate' to transfer the instance to the new computer. The screen below shows the status messages during a migration.



After the migration, you will find the virtual machine on the selected destination node.

26.9 Working with the console

The web interface of ArchivistaVM does not work directly with the instances. Instead, the web server rather executes the commands using utilities. These programs can also be accessed directly on the console.

26.9.1 QM Manager

With the QM Manager (activate on the console using `qm`), you can access the machine directly (without the web interface). The QM Manager is very powerful, hence the help texts which you will initially see if you access `qm` without further parameters.

```
qm <command> <vmid> [OPTIONS]
qm [create|set] <vmid>
--memory <MBYTES>      memory in MB (64 - 8192)
--smp <N>               set number of CPUs to <N>
--ostype NAME           specify OS type
--onboot [yes|no]       start at boot
--keyboard XX          set vnc keyboard layout
--cpuunits <num>       CPU weight for a VM
--name <text>           set a name for the VM
--description <text>   set VM description
--boot [a|c|d|n]       specify boot order
--bootdisk <disk>      enable booting from <disk>
--acpi (yes|no)         enable/disable ACPI
--kvm (yes|no)          enable/disable KVM
--tdf (yes|no)          enable/disable time drift fix
--localtime (yes|no)    set the RTC to local time
--vga (gd5446|vesa)     specify VGA type

--vlan[0-9u] MODEL=XX:XX:XX:XX:XX:XX[,MODEL=YY:YY:YY:YY:YY:YY]

--ide<N> [file=file,][media=d]
[ ,cyls=c,heads=h,secs=s[,trans=t]]
[ ,snapshot=on|off][,cache=on|off][,format=f]

--ide<N> <GBYTES>      create new disk
--format <format>       qcow|raw|raw2 => type of disk format
--ide<N> delete         delete disk
--cdrom <file>          is an alias for --ide2 <file>,media=cdrom

--scsi<N> [file=file,][media=d]
[ ,cyls=c,heads=h,secs=s[,trans=t]]
[ ,snapshot=on|off][,cache=on|off][,format=f]

--scsi<N> <GBYTES>      create new disk
--scsi<N> delete         delete disk

--virtio<N> [file=file,][media=d]
[ ,cyls=c,heads=h,secs=s[,trans=t]]
[ ,snapshot=on|off][,cache=on|off][,format=f]

--virtio<N> <GBYTES>      create new disk
--virtio<N> delete         delete disk

qm monitor <vmid>       connect to vm control monitor
qm start <vmid>          start vm
qm shutdown <vmid>      gracefully stop vm (send poweroff)
qm wait <vmid> [time]    wait until vm is stopped
qm stop <vmid>           kill vm (immediate stop)
qm reset <vmid>          reset vm (stop, start)
qm suspend <vmid>        suspend vm
qm resume <vmid>         resume vm
qm cad <vmid>            sendkey ctrl-alt-delete
qm destroy <vmid>        destroy vm (delete all files)
qm status <vmid>         shows the container status
```



```
qm cdrom <vmid> [<device>] <path> set cdrom path. <device is ide2 by default>
qm cdrom <vmid> [<device>] eject eject cdrom

qm unlink <vmid> <file> delete unused disk images
qm vncproxy <vmid> <ticket> open vnc proxy
qm vnc <vmid> start (X11) vncviewer (experimental)
qm showcmd <vmid> show command line (debug info)
qm list list all virtual machines

qm startall start all virtual machines (when onboot=1)
qm stopall [timeout] stop all virtual machines (default timeout is 3 minutes)
```

Here are a few commands. With `qm list` you will see a list of the currently defined instances.

```
avbox250a:~# qm list
VMID NAME          STATUS  MEM(MB)  BOOTDISK(GB) PID
101 test           stopped  512      32.00 0
104 avtest1        stopped  512      32.00 0
105 avtest2        stopped  512      32.00 0
106 avtest3        stopped  512      32.00 0
107 avvm           stopped  512      32.00 0
108 ArchivistaBox  stopped  512      32.00 0
109 test           running  512      32.00 10497
avbox250a:~#
```

With `qm showcmd 101` you will receive the command with which an instance is started (in our case the instance is 101).

```
avbox250a:~# qm showcmd 101
/usr/bin/kvm -monitor unix:/var/run/qemu-server/107.mon,server,nowait -vnc unix:/var/run/qemu-server/107.vnc,password -pidfile /var/run/qemu-server/107.pid -daemonize -usbdevice tablet -name avvm -smp 1 -vga cirrus -tdf -drive file=/var/lib/vz/images/107/vm-107-disk-1.qcow2,if=ide,index=1 -drive file=/var/lib/vz/images/107/vm-107-disk.qcow2,if=ide,index=0,boot=on -drive file=/var/lib/vz/template/iso/avtest2.iso,if=ide,index=2,media=cdrom -m 512 -net tap,vlan=0,ifname=vmtab107i0,script=/var/lib/qemu-server/bridge-vlan0 -net nic,vlan=0,model=rtl8139,macaddr=42:34:67:3F:92:48
avbox250a:~#
```

With `qm start 101` you can start an instance. Then you can use `qm list` to verify that the machine has actually started.

```
avbox250a:~# qm list
VMID NAME          STATUS  MEM(MB)  BOOTDISK(GB) PID
101 test           stopped  512      32.00 0
104 avtest1        stopped  512      32.00 0
105 avtest2        stopped  512      32.00 0
106 avtest3        stopped  512      32.00 0
107 avvm           stopped  512      32.00 0
108 ArchivistaBox  stopped  512      32.00 0
109 test           running  512      32.00 10497
avbox250a:~# qm start 101
avbox250a:~# qm list
VMID NAME          STATUS  MEM(MB)  BOOTDISK(GB) PID
101 test           running  512      32.00 2569
104 avtest1        stopped  512      32.00 0
105 avtest2        stopped  512      32.00 0
106 avtest3        stopped  512      32.00 0
107 avvm           stopped  512      32.00 0
108 ArchivistaBox  stopped  512      32.00 0
109 test           running  512      32.00 10497
avbox250a:~#
```

With `qm stop 101` you can perform a hard shutdown on the machine. This is not really recommended since, in certain cases, the (virtual) disk must subsequently be reorganised. With `qm shutdown 101` you can shut down Machine 101 in a controlled manner. It may be the case, however, that the operating system (instance) prevents this, in which case you will have to use `qm stop 101` nonetheless.

26.9.2 Data backup with vzdump

The data backup can also be executed using the console. For this, use `vzdump`. With `vzdump` you will receive an overview of all commands.

```
vzdump
Use of uninitialized value $opt_keepbackup in numeric eq (==) at /usr/sbin/vzdump line 891.
usage: /usr/sbin/vzdump OPTIONS [--all | VPSID]

--exclude VPSID          exclude VPSID (assumes --all)
--exclude-path REGEX     exclude certain files/directories
--stdexcludes            exclude temporary files and logs

--compress              compress dump file (gzip)
--dumpdir DIR            store resulting files in DIR
--extdev DEVICE          device to mount for backup
--keepbackup N           keep old backups (1-x)
--tmpdir DIR             store temporary files in DIR

--mailto EMAIL           send notification mail to EMAIL.
--quiet                 be quiet.
--stop                  stop/start VPS if running
--suspend               suspend/resume VPS when running
--snapshot              use LVM snapshot when running
--size MB               LVM snapshot size

--node CID              only run on pve cluster node CID
--lockwait MINUTES      maximal time to wait for the global lock
--stopwait MINUTES      maximal time to wait until a VM is stopped
--bwlimit KBPS          limit I/O bandwidth; KBytes per second

--restore FILENAME       restore FILENAME
```

26.9.3 Cluster management with pveca

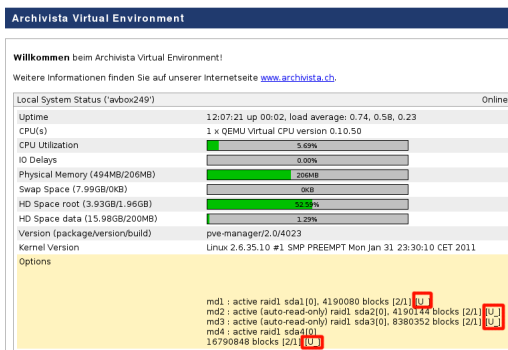
Cluster management via the console is described in 26.8.

26.10 Repairing hard disk networks

Hard disk networks are available on all ArchivistaBox 64 Bit versions. To check whether the network is running properly, simply log into ArchivistaVM.

26.10.1 Recognising a broken disk

If, after logging onto ArchivistaVM, you look under options, in the square brackets you should only see [UU] for two hard disks, [UUUU] for four or correspondingly more 'U's in the case of several hard disks. If, as on the following screen, you see a [U_], a [_U], or [UUU_] in the case of four disks, this means that a hard disk is defective and must be replaced.



➡: Don't worry. With good backups, this is not as serious as you might think. If you have not created any backups, it is advisable to do that first.

26.10.2 Identifying a defective hard drive

Once we know that a disk is defective, we need to determine which one it is. Log in to the console (with root privileges) and execute the command `hdparm /dev/md4`. Note on which disk the access indicator lamp does not light (about two to three seconds). This is the hard disk which you need to replace.

26.10.3 Restart server

It would be possible to replace the defective hard drive hot-plug-in. But then it can happen that the new inserted hard disk gets a new identification (in our example `/dev/sdb` can become `/dev/sdc`). Therefore a restart of the server is recommended.

➡ On some machines it can happen, that the boot order after a restart changes. In such a case just manually choose the boot drive. You can usually do this by pressing the 'ESC' or 'F8' key just after switching your computer on (check your computer manual).

26.10.4 Executing program `mdadm-repair.pl`

After rebooting, execute the utility `mdadm-repair.pl`. To do this type (with root privileges) the following two commands into the console:

```
cd /home/cvs /archivista/jobs/  
perl-mdadm repair.pl
```

You can now log in to ArchivistaVM, and in the overview, you should see under 'options' that the hard disk array has been rebuilt.

Options

md1 : active raid1 sdb1[2] sda1[0], 4190080 blocks [2/1] [UU]
=>.....] recovery = 53.3% (2235200/4190080) finish=1 smin speed=24808K/sec
md2 : active raid1 sdb2[2] sda2[0], 4190144 blocks [2/1] [UU] resync=DELAYED
md3 : active (auto-read-only) raid1 sda3[0], 8380352 blocks [2/1] [UU]
md4 : active raid1 sdb4[2] sda4[0], 16790848 blocks [2/1] [UU] resync=DELAYED

After some time (for large hard disk arrays the process can take several hours) you should once again see 'UU' in all positions (exception `/dev/md3` see below).

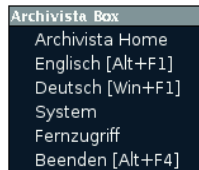
md1 : active raid1 sdb1[1] sda1[0], 4190080 blocks [2/2] [UU]
md2 : active raid1 sdb2[1] sda2[0], 4190144 blocks [2/2] [UU]
md3 : active (auto-read-only) raid1 sda3[0], 8380352 blocks [2/1] [U_]
md4 : active raid1 sdb4[1] sda4[0], 16790848 blocks [2/2] [UU]

➡: If `/dev/md3` is not aligned, then the swap partition has never before been used (marked green in the picture above). It will be synchronised later during the first access.

26.11 ArchivistaVM-Server

The ArchivistaVM module is available for every ArchivistaBox installed. Some of our customers operate ArchivistaVM as a server. In this case, it is referred to as ArchivistaVM-Server. It is available on our web shop shop.archivista.ch under ArchivistaVM-Servers for virtualisation, and includes significantly more power than is generally needed for a DMS.

For this reason there is an installation CD specifically for ArchivistaVM-Server. This is significantly smaller (approximately 330 MB) than the other versions. As a result, the ArchivistaVM installation CD only contains the software for virtualisation. Therefore, for such an installation, only those ArchivistaBox menu items which are not specific to the document management system (DMS) are available. These include, for example, remote maintenance and the settings for the system. The figure below shows the reduced ArchivistaBox menu:



➡: For all menu items present in ArchivistaVM-Server you can receive assistance under 25. Otherwise, no further distinction between the ArchivistaVM module and ArchivistaVM-Server will be made in this handbook.

Part VIII

ArchivistaERP

27 ArchivistaERP

27.1 Introduction

ArchivistaBox originated as a DMS and archiving software. Since the year 2008 ArchivistaBox contains an ERP (Enterprise Resource Planning) module. We at Archivista GmbH are convinced that the combination of an ERP with a DMS solution makes sense for various reasons.

The most important reasons are:

- Many clients ask for the possibility of an integration of their existing ERP Solution with ArchivistaBox. Thanks to ArchivistaERP there is an example and a test object.
- Today, box solutions are in demand more than ever. A simple web-based, ready-to-use ERP software does not exist. ArchivistaERP is therefore highly innovative.
- The number of webbased ERP solutions can be counted on one hand. Even OpenSource solutions are rare. ArchivistaERP is subject to the GPL and runs in entirely webbased mode.
- We ourselves (Archivista GmbH) needed a new ERP solution. During the evaluation phase we came across various products. The solution created originates from Frontaccounting (see www.frontaccounting.com) and was adapted to our needs so that we can work efficiently and speedily.

This is why ArchivistaERP came into being. In order that no misunderstandings arise let us mention what ArchivistaERP is not:

- ArchivistaERP is not a substitute for your existing ERP solution. The crucial point with an ERP solution is not the

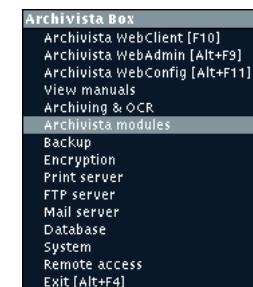
licencing cost but the implementation cost. If your ERP solution more or less does what you want it to do, then keep it.

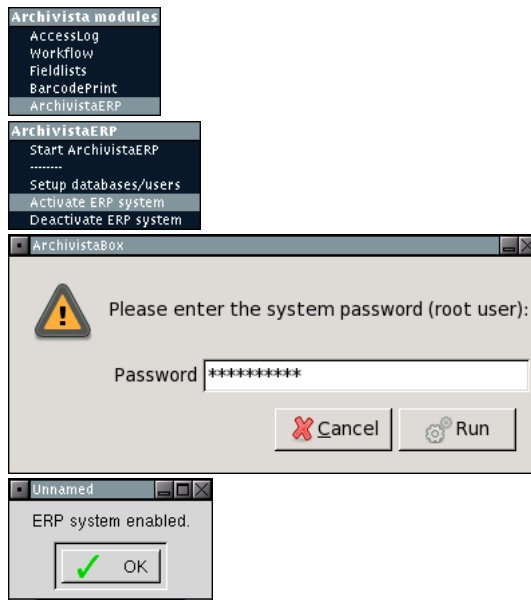
- ArchivistaERP is currently not geared towards medium and large enterprises.
- ArchivistaERP is one solution among many and ArchivistaBox supports all imaginable ERP solutions with a minimum of integration effort – totally independent of the ERP module discussed in this chapter.
- ArchivistaERP will be developed further in the future, i.e. certain adjustments will be made.

27.2 Activate ERP system

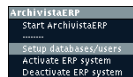
When ArchivistaBox is delivered, the ERP module is not active. For you to be able to work with the ERP module ArchivistaERP must be activated.

Click the right mouse button on the Archivista desktop. Select 'Modules' and go to ArchivistaERP:





After having activated the ERP solution, we can then create a database. Select in the same menu as above the following item:



The installation dialog appears. Enter the following information:

The screenshot shows the 'ArchivistaERP Installation Wizard' dialog. It has a title bar 'ArchivistaERP Installation Wizard' and a subtitle 'Welcome to the ArchivistaERP Installation Wizard.' Below is the Archivista logo. The main area contains fields for 'Database Name' (value: 'averpe'), 'Company Name' (value: 'Training Co.'), 'Username' (value: 'admin'), and 'Email' (value: 'webmaster@archivista.ch'). There are also password fields for 'Password' and 'Re-Password', all filled with asterisks. A dropdown for 'Accounting scheme' shows 'de_CH-new.sql'. At the bottom, there is a 'Please note' section with a license notice and an 'Install ArchivistaERP Accounting' button.

- Database name: name of database
- Password: root password of ArchivistaBox

- Company Name: name of your company (can be modified later)
- Database scheme: de_CH-new.sql, en_US-new.sql, en_US-demo.sql
- EMail: e-mail address of admin user
- Password/Re-Password: password of admin user (to be entered twice)

Click on 'Install ArchivistaERP'. Once all entries are completed, you will see the login screen:

The screenshot shows the 'ArchivistaERP 2009/I' login screen. It features the Archivista logo on the left and a login form on the right. The form has fields for 'User name', 'Password', and 'Company' (a dropdown menu showing 'Training Co.'). There is a 'Login' button and a link 'Please login here'.

You can now register yourself with 'admin' and the corresponding password. Click on 'Login'; you will get to the main view of ArchivistaERP.

The screenshot shows the main view of ArchivistaERP. It has a top navigation bar with tabs: Sales, Purchases, Items and Inventory, Manufacturing, Dimensions, Banking and General Ledger, and Setup. Below the navigation bar is a header area with 'Training Co. | localhost | Administrator' and a 'Logout' link. The main content area is divided into three sections: 'Transactions' (with links like Quote and Sales Order Entry, Direct Delivery, Direct Invoice, etc.), 'Inquiries and Reports' (with links like Sales Order Inquiry, Customer Transaction Inquiry, etc.), and 'Maintenance' (with links like Add and Manage Customers, Customer Branches, etc.). At the bottom, there is a status bar showing '05.01.2009 | 07:20' and 'ArchivistaERP 2009/I - Theme: default Archivista GmbH'.

The installation is complete. It makes sense to work through 27.3 first. Subsequently, you should specify settings: For more information go to 27.11.

27.3 First Steps

In this chapter we would like to take you step by step through the creation of an invoice. To begin with, we need the company data, at least one tax rate, a tax group, an article and a client.

27.3.1 Company Setup

After registration change to 'Setup' as well as 'Company Setup'. You can now enter your company data:

27.3.2 Taxes

Before we are able to create an article, we must ascertain that there is at least one entry for 'Tax Rate':

Create an entry along the lines of what you see below:

Besides the 'Tax Rate' we will need a 'Tax Group'. Please click on 'Tax Groups'.

Tip: The difference between 'Taxe Rate' and 'Tax Groups' lies in the fact that a 'Tax Group' may contain several 'Taxe Rates' and that one can specify for a 'Tax Group' if a delivery is subject to tax or not.

27.3.3 Managing items

Let us change to 'Items and Inventory'. Here an entry 'Item Tax Types' must be created next:

Proceed according to the illustration below:

Afterwards, we can set up the first article under 'Items':

Items

Enter a new item

Item Code: rigi

Name: ArchivistaBox Rigi

Description:

Image File (jpg): Browse

Category: Komponenten

Item Tax Type: MwSt

Item Type: Purchased

Units of Measure: Stück

Selling: ☒

Depending:

Barcode:

Weight:

GL Accounts

Sales Account: 3000 Verkauf

At the end, click on 'Insert new Item'. Now we have created the first article. What is still missing is that we have to assign a price to it.

Inventory Item Sales prices

Item: Komponenten - ArchivistaBox Rigi

The new price has been added.

Currency	Sales Type	Price	Factor	
CHF	Endkunden	800.00	0.00	Edit Delete

Currency: CHF - Franken

Sales Type: Endkunden

Price:

Factor:

Add/Update Price

Back

27.3.4 Adding a customer

To be able to start selling, we must create a customer account. Go to 'Sales' and then to 'Add and Manage Customers'. Fill in the form along the lines of what you see below:

Customers

Enter a new customer

Customer Name: Petermann Ltd

Address: Buchgasse 3
5400 Baden

Email:

GSTNo:

Customer's Currency: CHF - Franken

Sales Type/Price List: Endkunden

Dimension 1:

Discount Percent: 0.0 %

Prompt Payment Discount Percent: 0.0 %

Credit Limit: 1 000.00

Payment Terms: Zahlungen innerhalb 10 Tage

Credit Status: Gute Zahlungsmoral

Add New Customer

Back

Important: In ArchivistaERP the address is always entered in a memo field. At the same time the formatting of the address is defined.

In order to create a customer account, click on 'Add New Customer'. Additionally, for every client we must open a branch. You will find this under 'Add branches for this customer':

Customer Branches

Select a customer: Petermann Ltd

The selected customer does not have any branches. Please create at least one branch.

Name and Contact	GL Accounts
Branch: Petermann Ltd	Sales Account: 3000 Verkauf
Contact: Main Branch	Sales Discount Account: 3000 Verkauf
Phone: <input type="text"/>	Accounts Receivable Account: 1500 Debitoren
Fax: <input type="text"/>	Prompt Payment Discount Account: 3200 Rabatte, pünktliche Zahlungen
E-mail: <input type="text"/>	
Sales Person: Verkaufsperson	
Sales Area: CH	
Default Inventory Location: Standard	
Default Shipping Company: Default	
	Addresses
	Mailing Address: Buchgasse 3 5400 Baden
	Billing Address: Buchgasse 3 5400 Baden

Please note that every branch will need a new address where the the invoice and delivery can be entered. In our case, we simply copy the company's address details.

27.4 Creating an invoice

In order to create an invoice, choose 'Direct Invoice' in 'Sales':

To add an article, either enter the article code (here rigi) or use the drop down option to choose from the list. Finally click on 'Add Item'.

Direct Sales Invoice

Customer: Price List: Invoice Date:
Branch: Customer Discount: 0%
Reference:

Item Code	Item Description	Quantity	Unit	Price	Discount %	Total	
rgi	ArchivistaBox Rigi	1,00	St.	800,00	0,0	800,00	Edit Delete
	Komponenten - ArchivistaBox Rigi	1,00	St.	800,00	0,0	800,00	Add Item
						Total Excluding Shipping	800,00

Delivery Details

Deliver from:
Location:
Due Date:
Deliver To:
Address:

Contact:
Customer Reference:
Comments:
Shipping:
Charge:
Shipping Company:

The invoice will be created by clicking 'Place Invoice'. Due to the fact that some information must be updated internally, it may happen that the button 'Place Invoice' must be clicked twice.

Direct Sales Invoice

Invoice # 17 has been entered.

[View This Invoice](#)
[Print Sales Invoice](#)
[Enter a New Direct Invoice](#)
[Back](#)

The invoice is created and already booked too. It can be printed by pressing 'Print Sales Invoice' as a PDF. The result should look something like this:

Training Co.

Charge To:
Delivered To:

SALES ORDER

Invoice no.: 1 Cust no.: 1 Date: 31.12.2008

Shipping Company: Default Delivery Date: 10.01.2009
Your Ref.: Main Branch Our Ref.: Verkaufsperson
Your VAT no.: Our VAT No.:
Payment Terms: Zahlungen innerhalb 10 Tage Our Order No.:

Item Code	Item Description	Quantity	Unit	Price	Discount %	Total
rgi	ArchivistaBox Rigi	1,00	St.	800,00		800,00

➡ Please note that booked invoices cannot be deleted. Especially at the beginning this can happen, though. We recommend that you initially experiment with a sample tenant before the definite version is active. A new company can be created as described in 27.2.

27.5 Sales

In the sales section all quotes, orders and invoices are managed. Likewise, client address administration is done here.

Sales | Purchases | Items and Inventory | Manufacturing | Dimensions | Banking and General Ledger | Setup

Training Co. | localhost | Administrator [Logout](#)

Transactions

Quote and Sales Order Entry
Direct Delivery
Direct Invoice
Delivery Against Sales Orders
Invoice Against Sales Delivery
Template Delivery
Template Invoice
Customer Payments
Customer Credit Notes
Allocate Customer Payments or Credit Notes

Inquiries and Reports

Sales Order Inquiry
Customer Transaction Inquiry
Customer Allocation Inquiry
Customer and Sales Reports

Maintenance

Add and Manage Customers
Customer Branches
Sales Types
Sales Persons
Sales Areas
Credit Status Setup

05.01.2009 | 07:49

ArchivistaERP 2009! - Theme: default
Archivista GmbH

27.6 Purchases

Here purchase orders are managed. Supplier administration is part of this, too.

Sales | Purchases | Items and Inventory | Manufacturing | Dimensions | Banking and General Ledger | Setup

Training Co. | localhost | Administrator [Logout](#)

Transactions

Purchase Order Entry
Outstanding Purchase Orders Maintenance
Payments to Suppliers
Supplier Invoices
Supplier Credit Notes
Allocate Supplier Payments or Credit Notes

Inquiries and Reports

Purchase Orders Inquiry
Supplier Transaction Inquiry
Supplier Allocation Inquiry
Supplier and Purchasing Reports

Maintenance

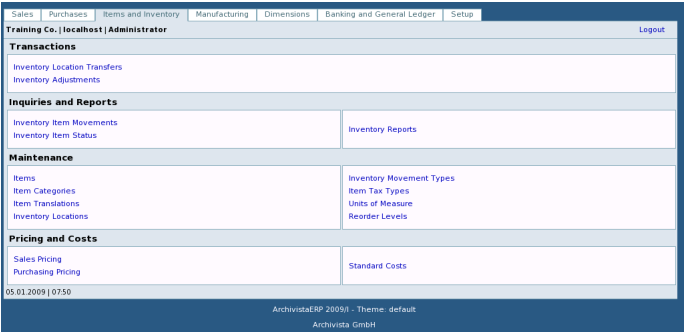
Suppliers

05.01.2009 | 07:49

ArchivistaERP 2009! - Theme: default
Archivista GmbH

27.7 Items and Inventory

In this menu we manage articles as well as the inventory.



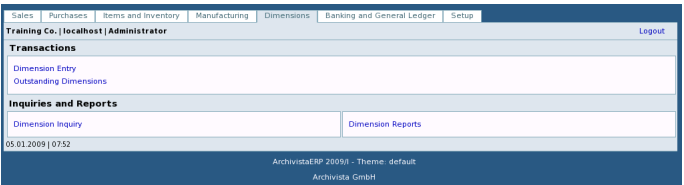
27.8 Manufacturing

Production planning can be carried out here.



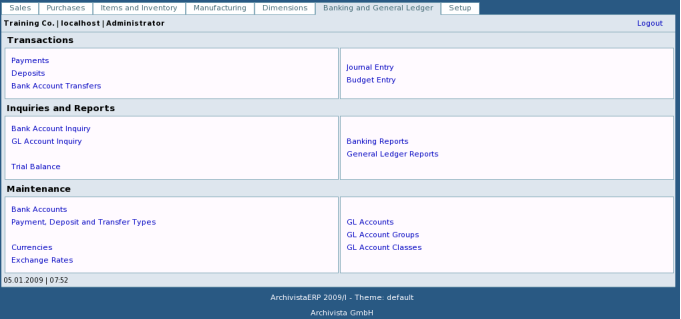
27.9 Dimensions

'Dimensions' means costing, cost centers and/or profit centers.



27.10 Banking and General Ledger

Here you will find the accounting functions as well as all options for payments and bank accounts.



27.11 Setup

In this main menu various options can be preset.



Part IX

Publishing

28 Publishing Edition III

28.1 Long-term considerations on documents

At the very beginning we should ask ourselves for which reasons we want to create an Archivist PublishingEdition. Let us assume that we want our documents to be available for a long period of time. 'Long-term' may mean lots of different things but let us think of a timespan between 20 and 50, or maybe even 100 years. This sounds like an infinite stretch of time in the world of computer technology but it is not in terms of paper documents.

Which documents will still be available at that point in time? Will we still be able to read an MS Word file properly in 20 years' time? Will the Internet still exist (admittedly a heretic question)? Where will Adobe's PDF format be? Will we still have the right to look at our own documents without a meter ticking away somewhere and intermittently making a message pop up that our account is running low?

28.2 The Schneider adventure

Please permit us to make a short digression here by relating an experience which had a lasting effect on our thinking. In 1983 a man walked into a shop and bought his first PC. The man was later to be managing director of Archivist GmbH and the PC was a Schneider CPC 464: it used a tape as storage medium. The salesperson in the retail shop who sold it said that the tape was a relatively unsafe medium but for short-term purposes it would be more than sufficient. Soon a floppy disk drive would come on the market and that would be extra safe and much faster too.

The storage capacity of 180 KBytes had great appeal but the Schneider computer was the only PC around with a 3" disk drive. Consequently, the shop assistant was asked if it was not a problem that Schneider had a 3" drive while all other home computers possessed either a 5.25 or a 3.5" drive? No problem, he answered, the 3" drive was very much on the upsurge, Amstrad was introducing it too.

The customer swallowed what the shop assistant said and bought the Schneider PC. Only much later did he realise that Schneider and Amstrad were the same and that his initial gut feeling proved right: no other 3" drive ever appeared on the market. The man stuck to his Schneider PC faithfully but the data of those years, i.e. 1983 to 85, could not be converted later on. It was his luck that he had printouts of all the files so that he could scan and archive them later. Without this experience Archivist might never have come into existence.

What are we to learn from this cautionary tale? In contrast to the 3" disk the tape has survived until today. And, the information could be saved because it existed as printouts. Does that mean that we have to produce printouts of all electronic files on our computer? The answer cannot be 'yes' here, also because we would not need a PublishingEdition then.

28.3 The XP adventure

The first version of the Publishing Edition was created within a client project in the year 2004. The task was to create a CD which could be started on a Windows PC without requiring any additional software. This solution was greatly loved by our clients

right up to the moment when with Windows XP the firewall was activated with SP1 by default.

Not that our solution did not run any more, but since our solution used local port 8088, the Windows firewall blocked this particular local port. In the final analysis this meant that for most users working with our Publishing Edition went from worry-free to hiccup-laden.

We could go into arguing that it makes little sense to shut the local ports in a local environment by default when at the same time (at least at the time) it was possible to write on the hard disk via the temporary directory. However, when you want to open a self-supporting archive, then such deliberations are hardly of interest to you. You simply want to work with your archive.

From a vendor's point of view it was a deplorable situation. We did make all the right choices with Web, MySQL and Windows, did we not? Bad luck when we propose but Windows disposes. Windows is not fond of Web technologies and this is not our fault. But even this will be of little interest to you when you are trying to make archived data available.

As long-term reaction to this (and other) experiences ArchivistaBox has come into being: a solution which on the one hand keeps favouring Web technologies and on the other can be deployed as stand-alone, embedded box solution to be at all events prepared for future imponderabilities.

Until a few months ago using an ArchivistaBox required that the software be set up and the data to be archived needed to be added. Because of this, ArchivistaBox could not produce self-supporting archive CDs. It was then decided to enable ArchivistaBox to create self-supporting CDs anytime from a running system. These CDs were to serve a double purpose: for one they should be liveCDs viewable as live archives, and for another they should serve as

installation CDs whereby an archive could be put back onto the server and where new data could be added.

We can only be grateful that at that time the Windows firewall shut the local ports because it gave us the possibility to make an effort towards a solution which is in every direction much more open than the previous Publishing Edition.

With the new solution (which is - it goes without saying - 100 percent OpenSource) no Window license is needed nor do we have to list any version numbers or ServicePacks as prerequisites. All we need is: a) a computer with 256 MByte RAM or b) a virtual box with 256 MByte RAM. Everything else is handled by the ArchivistaBox: and suddenly publishing archives is easy.

We are of the opinion that the Roman Numeral II in the chapter heading is justified. We wish you and us much success and pleasure with the new Publishing Edition.

28.4 Why source formats are not suitable

How can we ensure that our documents keep well and are safe for long periods of time? Why are source formats (e.g. MS Word files) unsuitable and why should we also convert PDF files although they carry 'portable document format' in their name? Simply because we need powerful programs to display these files. All of them contain characters, more or less exotic fonts, images, and sometimes even bits of executable code, that need to be put together on the fly and with the help of considerable processing power before we can view our file on screen.

Keeping documents can be massively simplified if we create an image file out of our source format. The image file resembles a photograph in that it freezes the look of a file at a particular moment in time. Display on screen is much simpler too. At Archivista Gmbh we speak of 'virtual printing' because converting source for-

mat files into image files entails nothing more than printing them to a file and rasterising them. Finally, they end up as image files in the Archivista database.

In the following I would like to explain the abovementioned process step by step and with the help of Archivista. It saves us the task of having to print out our documents, file and re-scan them again. Of course, you can also publish documents that exist only as hard copies to begin with: these you scan first. However, the following tutorial is dedicated exclusively to documents already existing in digital form on your computer.

The points made in this section are valid for all archives whether they were created with the Archivista PublishingEdition or not. We should be aware of them when we later encounter problems during the import of documents. You may safely assume that if certain files give you problems when you convert them today, they would generate immensely more hassle if you tackled them in, say, ten years.

Have fun with the following tutorial!

28.5 Preliminary note

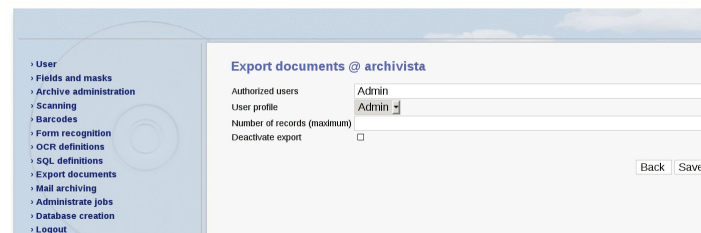
In order to publish Archivista databases you need an ArchivistaBox installation. How to install ArchivistaBox is described in 2. As soon as this is done you can add documents. Afterward you can continue with the following menu item, see 28.6.1.

Before you can create your first self-supporting archive you have to a) have the ArchivistaBox ISO file at hand and b) create an empty database.

28.6 Publishing self-supporting archives

28.6.1 Enabling file export in WebAdmin

After the installation, publishing must first be enabled in WebAdmin.



Those users who have the permission to publish archives must be entered in "Export of files" (here "Admin"). It is important to note that users entered here can duplicate the archives at the push of a button and if this is used in an uncontrolled fashion any ArchivistaBox will run out of space relatively quickly. For these reasons, users must be explicitly enabled. Run the command.

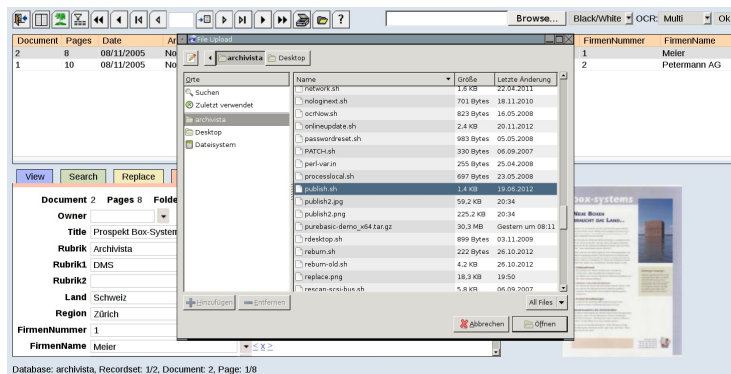
28.6.2 Provide ArchivistaBox ISO

In order to create the CD (ISO file) later, the ISO file from ArchivistaDMS must be available for the installed instance under:

```
/var/lib/vz/template/iso/archivista_cd1.iso
```

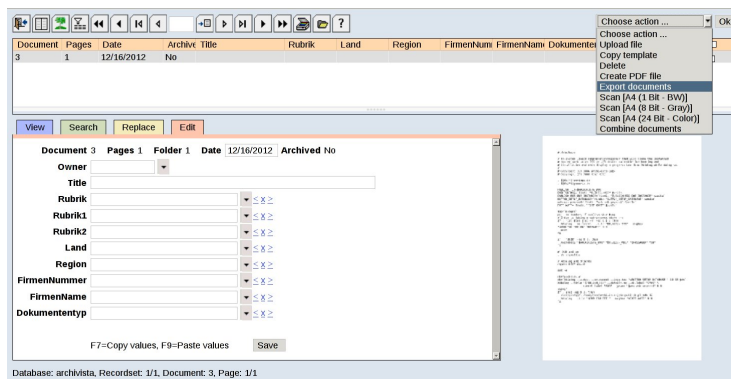
28.6.3 Uploading documents in WebClient

Now switch to the WebClient, and upload several documents to follow our example. In our example, we will use the script "publish.sh" directly on the ArchivistaBox.



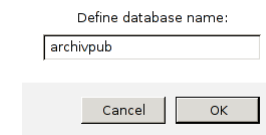
Any other documents (including MS Office files) may also be placed in the archive instead of this file. It should be borne in mind that a maximum of 4 GB of data can be published per archive because otherwise a new DVD cannot be created.

28.6.4 Selecting the documents for publishing



Before publishing can be started, the required documents must be selected using the "Search" function. Next, go to "Edit" and select "Export documents".

28.6.5 Exporting files in WebClient



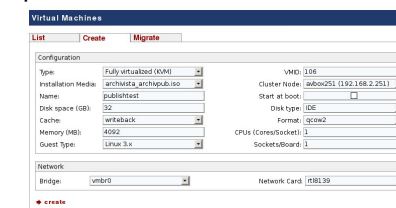
Now, enter the name of the new database (in this case, archivpub) to export the current selection to this database.



In order to automatically create an ArchivistaBox CD (ISO file) after you have exported the selection to the required database, confirm the following query with "OK."

28.6.6 Testing the published ArchivistaVM archive

Now go to ArchivistaVM (it is included with every ArchivistaBox). Just as a quick reminder, log out of the WebClient, then click on "Home" and login to ArchivistaVM using "root" and your root password.

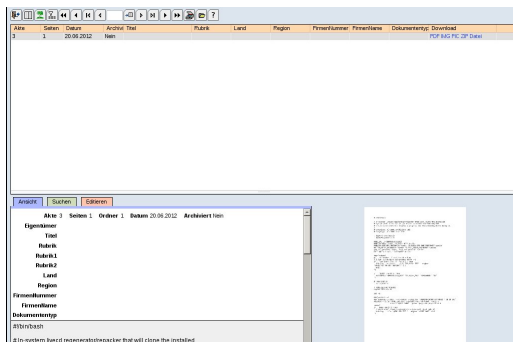


The published CD can now be tested in a virtualised environment. For this purpose, a virtualised instance must be set up in ArchivistaVM. Select the created CD with the name archivista_archivpub.iso (or the previously used name) under "Installation media." Furthermore, at least 4 GB memory must be available in RAM mode. To be more exact, there needs

to be 3 GB available in the RAM mode plus roughly the size of the created ISO file.

Important: if the ISO file does not exist, you must check "view system files" in WebConfig to see why the ISO file has not (yet) been created. In our example, it takes 20 to 30 seconds to create the ISO file. In the case of larger archives, this should not exceed 1 to 2 minutes.

Once the virtualised instance has been created, the CD can be started. The published ArchivistaBox created starts directly with 'ram ramonly', i.e. the ArchivistaBox will be started and set up in RAM.



Please note: if the published archive cannot be started in RAM mode, a normal installation can be performed on the first screen using the "linux" option.

The ISO file created can then be burned to a CD/DVD. If you want to do this directly from the ArchivistaBox, a terminal window must first be opened on the ArchivistaBox desktop (ALT-F7). Subsequently, use the "su" command to change to the root user. The command "cdrecord -scanbus" can be used to search for a burner. Next, burn the CD/DVD.

```
cd /var/lib/vz/template/iso
```

```
cdrecord archivista_archivpub.iso dev=1,0,0 speed=4
```

It is important to specify the correct device number. The number found with "cdrecord -scanbus" (e.g. dev=7,0,0) should be entered instead of dev=1,0,0.

Conclusion: with ArchivistaDMS, self-supporting archives can be created within minutes directly in the WebClient and tested in a virtualised environment directly on the same machine. The above example can be run through in the RAM, if you have 4 GB (or better, 8 GB) RAM available.

28.7 A few tips on working with documents

In the following we would like to give you a number of tips of how to get the documents into the ArchivistaBox in the first place (before publishing them in a second step).

28.7.1 Working with paper documents

Documents you receive as so-called hard copies must be scanned. Before starting this time consuming process you should make sure that no PDF file exists. As a rule you should be able to proceed much faster with electronically available documents if you possess the needed applications and the Archivista printer is set up properly on the ArchivistaBox.

300dpi yield good results when scanning. For the sake of velocity colour documents can be captured with lower resolution, however, it must be borne in mind that the text recognition might suffer accordingly.

For larger volumes we recommend to acquire a scanner with automatic document feeder if it is possible to split the documents into individual pages. Books often defy this procedure. For books a flatbed scanner may be used or a digital camera. Please note that with digital cameras lighting is essential – artificial light might give problems – as well as a tripod to hold it firmly in position. For

A4 pages a camera should possess at least a resolution of 6 mio pixels.

28.7.2 Working with electronic files

To transfer individual files to the archive we proceed in the following manner. We start up the application and open the particular file in question. Then, we print it using the printer named 'Archivista.'

➡ For information on how to set up a printer on the ArchivistaBox to print documents into the archive please see below.

28.7.3 Working with PDF files

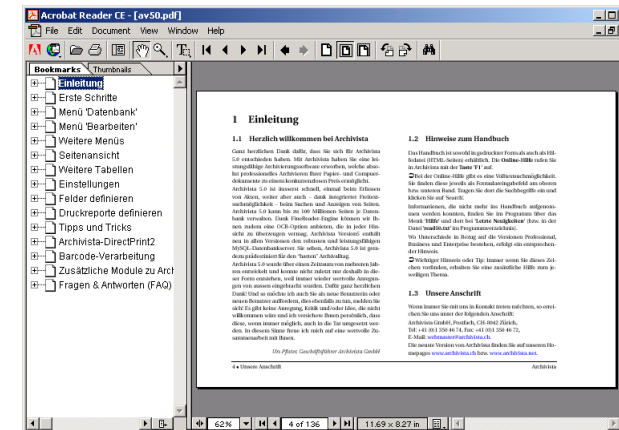
In the following we would like to discuss a few peculiarities of working with PDF files. To begin with let us bear in mind that the acronym 'PDF,' short for Adobe's 'Portable Document Format,' comprises a large number of different versions of the format. This diversity of formats is further increased by innumerable third-party products which emulate the Adobe Acrobat Writer and produce PDF files of good or lesser quality.

Unfortunately, version 6.0 of the Acrobat Reader is not as lean an application as earlier versions were. And it is true that sooner or later, i.e. when the latest version of the Acrobat Writer has established itself, you will have to use the most recent version of the Acrobat Reader. However, the principle we want to draw attention to here has remained more or less the same from the Acrobat Reader version 4.0 onwards. An older version may possibly serve our purpose better.

28.7.3.1 Difficult PDF files

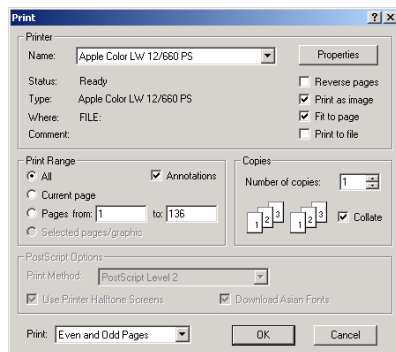
PDF files often create difficulties. In the Acrobat Reader (but also in other products) there is an option to print these PDF files

directly as image files. In the following an example: Open any PDF file you like (or the Archivista manual as in the example below).



Go to the menu 'Print...' (do not use the printing icon in the toolbar) and have a look at the printing options. For our purposes it is essential that a) you use the Archivista printer and b) you tick the option 'Print as image.' With the latter option each page is rasterized as image in the first place (no attempt at recognising fonts is made, for example). This way rasterising takes a bit longer but the fonts are interpreted more clearly than when we rasterize without this option.

Depending on the Acrobat Reader version you use you may have to search for this option in different places, as not all Acrobat versions are the same. If you do not find it straight away, please consult the Acrobat Reader manual.



Now print the file to the Archivista printer as usual and you will get images of high quality, especially with color publications.

28.7.4 Working with web pages

Archiving web pages is often a troublesome task. The reasons are manifold and we cannot go into all of them. We can discuss only a few major obstacles and how to remove them.

First of all we should not forget that the major hindrance to simply saving a web page as file onto our hard disk is that web pages normally consist of several files. Sometimes web pages are put together dynamically as when its contents are retrieved from a database.

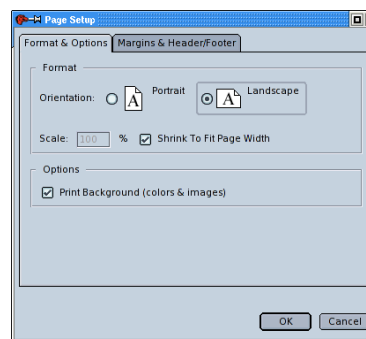
The easiest solution is if the web page we want to archive possesses a link to a print version. Some web pages offer this possibility because printing is explicitly desired.

If there is no print version available we should make sure that in terms of browsers there is not only the Internet Explorer installed but also Mozilla or Netscape (both rely on the same technology). Normally we get better results with Mozilla but in some cases Microsoft's Internet Explorer yields a more satisfying print output. If both fail, try a third type of browser (like Opera): it might give you the result you want.

Let us take a look at a web page in the print preview. It is not what we expect because the background is turned off and the preview's orientation is 'portrait.'

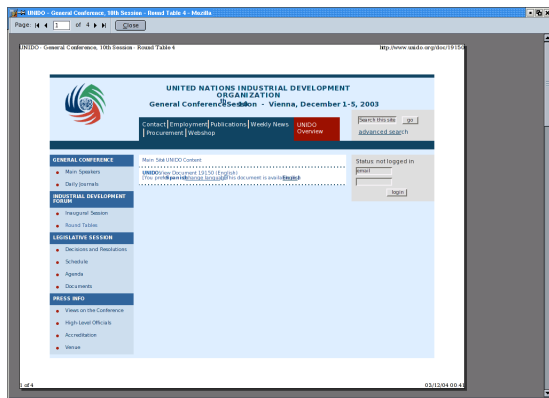


In 'Page Setup' we can change these values:

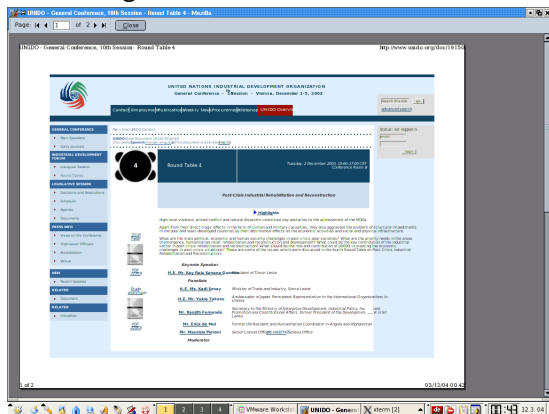


You find these options either directly in the 'Print preview' or, depending on the version of the browser you are using, in menu 'File.'

After these adjustments the page will look much better although it may happen that the contents have been miraculously moved to page two.



We can either start printing from page two or change the zoom factor to get a better result.



As soon as the print preview looks acceptable we print the page with the Archivista printer. The subsequent import of the page should not create any problems.

Part X

Internals

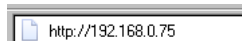
29 External SQL calls (API - WebClient)

For automated access to a particular database via WebClient we advise to follow the procedure and the conventions below.

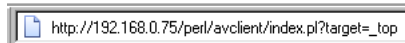
In principle, access via WebClient (e.g. to an ArchivistaBox) is defined by its IP address (or DNS name resolution). To access a specific database with a specific user, for example, it is possible to simply add some information to the IP address.

29.1 External login

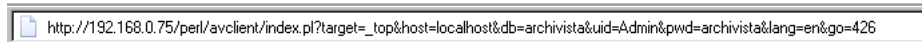
When the address of an ArchivistaBox is entered in the browser,



and 'Enter' is pressed, the IP address will automatically be complemented with a few internal parameters.



To this string of characters some other parameters can now be added in order to log in directly to a certain depth of the archive. The following, complete entry expresses the way with which the user reaches the view of the archive he needs or wants without him/her having to go through the individual steps manually. She/he wants 'localhost' for host, 'archivista' for database (with user 'Admin' who possesses password 'archivista'), and she/he wants an English layout but only document 426 to appear in the list.



The conventions are:

host	host computer
db	database
uid	user name
pwd	password
lang	language; the entry is en for English and de for German
go	go directly to document; here a document number is required

The order of the expressions between the & signs does not play a role. You can, for example, enter the password before the host. However, attention must be paid to making all necessary specifications: host and database are compulsory. 'lang' for language governs the layout, which means that for the sake of display it is compulsory, too. 'go' is not compulsory.

29.2 Extended access

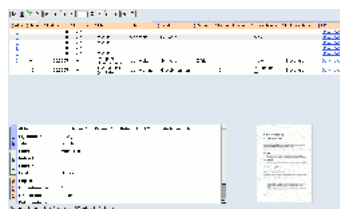
The ArchivistaBox 2008/III (and higher releases) can be used to control the entire WebClient externally as well as carry out registration and file selection automatically. This means that it is now possible to control the WebClient 'remotely' via another (web) application. Firstly, this application has to connect to the WebClient (see 29.1). The application can then call all the functions of the WebClient that are available in 'manual' operation.

All editing functions are available here. Files can be deleted, scanning is possible and indexing can be automated. In order to gain a better understanding of external access, it is important that you are aware of the various forms (main/page view) and views (display, search, edit) since not all of the commands are available in all views.

The WebClient has two forms (views): the main view (`go_main`) and the page view (`go_page`). There are three different modes in these two forms (statuses): display (`go_view`), search (`go_search`) and edit (`go_edit`) documents. Before the individual commands are described, these different form views are shown here in brief form:

29.2.1 Forms

29.2.1.1 Main view



You can use `go_main` to switch to the main view at any time. Here, the system displays a table with the data records that are currently activated in the overview as well as the active data record in the detail view and a thumbnail.

29.2.1.2 Page view



You can use `go_page` to activate the page view. This allows you to display pages in enlarged form. You are also provided with functions for rotating pages and carrying out post processing activities.

29.2.1.3 Switching between the page view and the main view

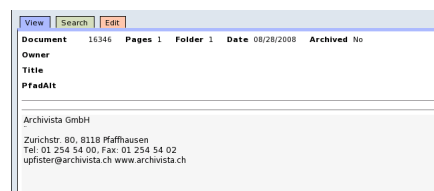
You can use `go_pageswitch` to switch between the main view and the page view.

29.2.1.4 Switching to photo mode

In photo mode, the pages are displayed as small thumbnails. Thus, the command for this mode is `go_thumbs`.

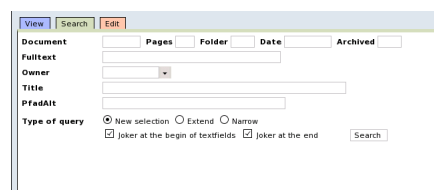
29.2.2 Modes (statuses)

29.2.2.1 Display



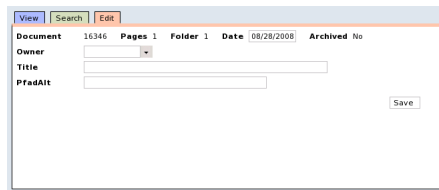
You can use `go_view` to switch to display mode. This displays the values that are currently saved in the database.

29.2.2.2 Search



You use `go_search` to activate the search form. You use `go_query` to trigger a search process and the fields are to be displayed with `fld_Field1=Value1&fld_Field2=Value2`.

29.2.2.3 Edit



You use `go_edit` to access editing mode. You can use `go_update` to edit the field values of actual files by entering the fields with `fld_Field1=Value1&fld_Field2=Value2`. However, this is only possible if editing rights exist in the respective file.

29.2.3 Logging in and logging out

29.2.3.1 Login

The login procedure is described earlier in the handbook under 29.1. Once you have logged in, the parameters are not required until the end of the session (i.e. when you logout).

29.2.3.2 Logout

You use `go_logout` to end the current session. The system then displays the form for logging in.

29.2.4 Navigation within files

29.2.4.1 Select all data records

You use `go_all` to display all files for the active session. This is also the case when the session is started and the first files are displayed.

29.2.4.2 Display previous data records

You can use `go_docs_prev` to display the previous data records in the selection. This is only useful if the first file is not already activated.

29.2.4.3 Previous data record

You use `go_doc_prev` to select the previous data record.

29.2.4.4 Selecting a data record

You can use `go_select` and `selectnr=xx&selectpage=yy` to activate a specific file and page from the **current selection**. You have the option of specifying a page but if you do not, the first page is always activated. This function is used primarily if you want to display a specific document or page after a search query (`go_query`).

The following example describes the selection of page 50 from document 4 (0 is the first document) from the **current selection**.

```
http://localhost/perl/avclient/index.pl?
go_select&selectnr=3&selectpage=50
```

➡ The above example uses the address 'localhost'. This only applies if you are working locally on an ArchivistaBox. In all other cases, 'localhost' must be replaced with the IP address or the name of the computer (domain).

29.2.4.5 Next data record

You use `go_doc_next` to activate the next file.

29.2.4.6 Display next data records

You use `go_docs_next` to display the subsequent files within the selection.

29.2.5 Navigation within pages

29.2.5.1 Display first page in document

You use `go_page_first` to select the first page in the current document.

29.2.5.2 Previous page

You use `go_page_prev` to select the previous page in the current document.

29.2.5.3 Jump to a specific page

You use `go_page_jump` to activate a specific page directly.

29.2.5.4 Subsequent page

You use `go_page_next` to select the subsequent page in the current document.

29.2.5.5 Display last page

You use `go_page_last` to select the last page in the current document.

29.2.6 Search in archives

29.2.6.1 Trigger search queries

You use `go_query` to trigger the search process.

To do so, you have to specify a field with `fld_Fieldname=Value` (or multiple fields) and/or a full text query with `fulltext=value`.

You use `searchmode=xxx` to determine the type of query. You have the following options for this:

- `searchmode=new` for a completely new query
- `searchmode=or` for extending the search query (or)
- `searchmode=and` for extending the search query (and)

Throughout the entire query, you can specify whether you want the system to use wildcards for the search or not. If you use `jokerstart=1`, a wildcard is used at the start of the field and if you use `jokerend=1`, a wildcard is used at the end of the field. The required values merely have to be somewhere within the field you require for both of these options.

The following example illustrates the search process:

```
http://localhost/perl/avclient/index.pl?
go_query&searchmode=new&
fld_title=Meier&fulltext=vacuumcleaner&
jokerstart=1&jokerend=1
```

When you start a new search query, this query searches the title field for 'Meier' with the additional prerequisite that the word 'vacuumcleaner' also has to occur in the full text. Wildcards are used at the start and end of fields meaning that all 'Meierhans' and 'Hansmeier' entries are also found.

29.2.6.2 Sorting search queries

You can use the `go_order_desc` and `go_order_asc` commands to sort a query in descending or ascending order for a specific field. To do so, you also have to specify the desired field accordingly with `orderfield=Fieldname`.

For example:

```
http://localhost/perl/avclient/index.pl?
go_order_asc&orderfield=Date
```

When you use this command, the current selection is sorted in accordance with the 'Date' field.

29.2.7 Editing files

By using `go_update` and entering fields (e.g. `fld_Title=Meier`), it is possible to carry out post processing for the files. It is only possible to edit fields if the corresponding rights exist in the required file. If this is the case, editing is carried out without any further queries from the system.

For example:

```
http://localhost/perl/avclient/index.pl?
go_update&fld_Title=Meier&fld_lNo=1001
```

This command sets the 'Title' field to 'Meier' and '1001' is entered in the 'No' field.

29.2.8 Manipulating the display

29.2.8.1 Rotate left

You use `go_rotate_left` to rotate the current page to the left.

29.2.8.2 Rotate by 180 degrees

You use `go_rotate_180` to rotate the current page by 180 degrees, that is, to turn it upside down.

29.2.8.3 Rotate right

You use `go_rotate_right` to rotate the current page to the right.

29.2.8.4 Zooming in and out on pages

➡ The following functions are only available in the page view since you can only zoom in or out on a page there.

You can use `go_zoom&zoom=x` to change the display continuously. Values between 0.1 and 1 (10 to 100 percent) are possible for `x`.

You use `go_zoom&zoom=0` to display the page in accordance with the available screen area.

You can also use `go_zoom_in` and `go_zoom_out` to zoom in or out on the document depending on the current view.

29.2.9 Further commands (action)

Under normal circumstances, some commands can only be accessed in the WebClient if you activate editing for a page or file beforehand. You can use `go_action&action=xxx` to access these commands, regardless of whether you activate editing beforehand. The only condition is that the file can be changed, i.e. the relevant user is authorised to do so.

29.2.9.1 Deleting files

You can use `go_action&action=delete` to delete the current file.

You can delete multiple files directly by adding `seldocs=x,y,z` and using the relevant file numbers in place of `x`, `y` and `z`. You can delete a maximum of 16 files directly.

For example:

```
http://localhost/perl/avclient/index.pl?
go_action&action=delete&seldocs=1,4
```

Files 1 and 4 are deleted in the active database.

➡ Deletion is carried out without any further queries from the system. However, files to be deleted must not be blocked (undergoing processing) or already archived. Files such as these cannot be deleted.

29.2.9.2 Publishing files

You can use `go_action&action=publish&owner=xxx` to publish archives.

➡ For 'owner', you have to enter the desired owner or the relevant group. If you want to publish the file for all users, you can do so by using `owner=[ALL]`. You can also use `seldocs=x,y,z` to publish multiple files directly.

29.2.9.3 Undoing published archives

Publishing can also be undone. You can use `go_action&action=unpublish` to do this. Once again, you can use `seldocs=x,y,z` to process multiple files simultaneously.

29.2.9.4 Combining files

You can use `go_action&action=combine` to combine two or more files (i.e. they are merged to form one file). You must enter `seldocs=x,y` here so that two files can be transferred to one file. The target file in this case is the active file.

29.2.9.5 Scanning files

You can use `go_scan` to scan a new file. A prerequisite for this is that a scanner is connected to the respective ArchivistaBox.

In addition, you can use `append_to=xxx` to carry out a scan in an existing file. You enter the desired file number in place of 'xxx'.

You use `scandef=y` to specify the desired scan definition.

Furthermore, you can also include field information with a file during scanning. You can do this by using `meta=Fieldname:Value`. Here, multiple fields can be addressed with multiple `meta` blocks.

The following example is intended to clarify the various possibilities during scanning:

```
http://localhost/perl/avclient/index.pl?
go_scan&append_to=44&scandef=3&
meta=Titel:Meier&meta=OrderNo:1001
```

If the above command is submitted as a whole line under 'localhost', the scanning process is triggered and the scanned pages are added to file 44. The third scan definition is used in this instance and the file then receives the value 'Meier' in the 'Title' (you have to use the german name 'Titel') field and the value '1001' in the 'OrderNo' field.

29.2.9.6 Delete page

You can use `go_action&action=deletepage` to delete a single page.

29.2.9.7 Saving changes to a page

You can use `go_action&action=savepage` to save a page that has been, for example, rotated.

29.2.9.8 Changing the OCR definition for a page

You can use `go_action&action=ocrpage!x` to assign a new OCR definition to a page. This ensures that this page is processed with the newly assigned OCR definition the next time text recognition (OCR) is called.

29.2.9.9 Setting OCR to complete for a page

You use `go_action&action=ocrdone` to set a page to complete for the OCR recognition. In other words, it is not used for the text recognition.

29.2.9.10 Excluding a page from OCR

You use `go_action&action=ocrexclude` to exclude the current page from any text recognition.

➡ You should use this option if certain pages within files only contain images since the OCR text recognition does not provide usable text and does not work quickly in the case of images.

29.2.9.11 Transferring the OCR definition to the file

You use `go_action&action=ocrselectdoc` to transfer the OCR setting of the current page to the entire file.

29.2.9.12 Transferring the OCR definition to the selection

You use `go_action&action=ocrselectdocall` to transfer the OCR definition of the current page to all of the files displayed.

29.2.10 Printing pages

You can use `go_print&print_from=x&print_to=y` to print a file. However, this is only possible if a postscript printer is installed on the respective ArchivistaBox and is activated for the users in Archivista WebAdmin.

29.2.11 Output format

Normally, when the files are accessed externally, the normal view of the WebClient is displayed. In order to make it as simple as possible to call information, a few output formats are provided so that specific information can be obtained directly. These display modes are outlined in the following.

29.2.11.1 Obtain page text

You can use `go_text` to obtain the current page of a file as a text file.

29.2.11.2 Obtain PDF file

You can use `go_pdf` to obtain the current file as a PDF file.

➡ Note: depending on the size of the file, it may take some time to obtain a file as a PDF file since the whole PDF file has to be created and delivered first.

29.2.11.3 Display an image file

You can use `go_image_doc_page_width_height` to call a page directly as an image. Here, `doc` represents the file number, `page` represents the page number, `width` represents the maximum width and `height` represents the maximum height.

For example:

```
http://localhost/perl/avclient/index.pl?
go_image_2108_2_1000_1000
```

Page 2 of file 2108 is displayed. The image is displayed with a maximum height and width of 1000 pixels.

29.2.11.4 Display Ajax list

You can use `go_ajax&ajaxfield=Fieldname&ajaxval=xx` if you want the system to read the values from Ajax fields (these are fields in the WebClient for which a window with corresponding values is opened when you perform a search or carry out editing).

For example:

```
http://localhost/perl/avclient/index.pl?
go_ajax&ajaxfield=Owner&ajaxval=a
```

In this example, all available owners for a file are returned. This information could, for example, be displayed as follows:

```
<ul><li>Admin</li>
<li>Admin1</li>
```

29.2.12 Return values as an HTML form

Normally, when the files are accessed externally, the normal view of the WebClient is displayed. Since this display contains extremely complex forms, you can use `frm_Fieldname=1` to ensure that only specific fields are shown in a simplified display.

For example:

```
http://localhost/perl/avclient/index.pl?
go_query&searchmode=new&fld_pages=4&
frm_filenummer=1&frm_pages=1
```

All files that contain precisely four pages are to be called. The 'Document number' and 'Pages' fields are to be displayed. Unsurprisingly, 'Pages' always contains the value '4' in the form:

```
<input type="hidden" name=" filenummer _1" value="1251">
<input type="hidden" name=" pages _1" value="4">
<input type="hidden" name=" filenummer _2" value="1229">
<input type="hidden" name=" pages _2" value="4">
<input type="hidden" name=" filenummer _3" value="1228">
<input type="hidden" name=" pages _3" value="4">
...
```

29.2.13 Export structural data as ANSI file

To export a query result to a spreadsheet program you can create an ANSI list directly in the browser with `frm_Laufnummer=2`.

29.2.14 Final example

To round off, the following program example illustrates a simple session with the WebClient in the Perl language:

```
use strict;
use LWP::UserAgent;
use HTTP::Cookies;
my $server = "http://localhost/perl/avclient/index.pl";
```

```
my $connect = "?host=localhost&db=archivista&uid=Admin&pwd=archivista";
my $www = LWP::UserAgent->new; # new www session
$www->cookie_jar(HTTP::Cookies->new('file'=>'/tmp/cookies.lwp','autosave'=>1));
my $res = $www->get("$server$connect"); # connect to webclient
if ($res->is_success) {
    if ($res->content) { # we got login
        my $cmd = "?go_query&fld_Seiten=3"; # search for 3 pages
        $res = $www->get("$server$cmd");
        my $cmd = "?go_doc_next";
        for (my $c=1;$c<=18;$c++) { # navigate 18 times to next doc
            $res = $www->get("$server$cmd");
        }
    }
}
```

The above example creates a connection with the WebClient, executes a search query in which all files with three pages are activated and then navigates through 18 files. Furthermore, when a script is called externally, it is important that the 'cookies' are saved. This is the only way in which further 'commands' can be transferred to the WebClient.

29.2.15 Transferring files to the archive

Since it is possible in the WebClient to transfer a file (PDF or graphic file) to the archive server, it is also possible to do this via external control. In principle, the `go_action&action=upload` command can be used for this. However, when you upload data, it is necessary to switch to 'POST' mode during communication with the WebClient (this is different from the GET mode that is normally used).

During the upload, you can include the `uploadocr=x` and `uploadbits=y` options in order to determine OCR definitions as well as scanning. Alternatively, you can also determine a scan definition containing the processing parameters directly during the upload. You use `uploaddef=x` to select a specific scan definition, although the first scan definition is activated with 0.

The following example demonstrates uploading a file to the web server in the Perl programming language. The comments should

be of assistance even if you carry out the upload in a different programming language.

```
#!/usr/bin/perl

# wcupload.pl -> demo script for uploading a document via web client
# (c) 2008 by Archivista GmbH, Urs Pfister

use strict;
use LWP::UserAgent; # we work with UserAgent (our batch web browser)
use HTTP::Cookies; # we need to work with cookies
use HTTP::Request::Common qw(POST); # the post method must be imported

my $fin = "/home/archivista/documentation_de.pdf"; # demo doc (our manual)
my $pages = "1-5"; # pages to extract (so we don't wait too long)
my $fout = "/tmp/eins.pdf"; # the finally file we want to import
if (!-e $fout) { # create file with pdftk if it does not already exist
    system("pdftk $fin cat $pages output $fout");
}

# server we use (link to webclient)
my $server = "http://localhost/perl/avclient/index.pl";
# connection string (host,db,user,password)
my $connect = "?host=localhost&db=archivista&uid=Admin&pwd=archivista";

my $www = LWP::UserAgent->new; # new www session
# save the cookie for the current session
$www->cookie_jar(HTTP::Cookies->new('file'=>'/tmp/cookies.lwp','autosave'=>1));
my $res = $www->get("$server$connect"); # connect to webclient
if ($res->is_success) {
    if ($res->content) { # we got login
        # now upload a file, we use request method with POST
        my $res = $www->request(POST "$server",
            Content_Type => 'form-data', # multipart/form-data
            Content => [ # structure for our file
                MAX_FILE_SIZE => 134217728, # max. size (WebClient won't accept more)
                upload => [ $fout, $fout ], # file to upload, file name to use
                go => 'go_action', # we need to call the go_action command
                action => 'upload', # inside of go_action we need to use upload
                uploadbits => 1, # 1=black/white, 8=gray, 24=color
                uploadocr => 27, # the desired ocr def (1-x, 27 does mean: no ocr)
                meta => "Titel:450", # filling in some meta keys is no problem
            ]
        );
        if ($res->is_success) { # if we got a success, file is uploaded
            print "file $fout uploaded\n";
        }
    }
}
```

➡ **Note:** The above example runs in every ArchivistaBox. In principle, however, the WebClient can be addressed by any computer and it is even possible to work with encrypted access (HTTPS). You can find the program on the ArchivistaBox

CD under /home/cvs/archivista/jobs under the name wcupload.pl.

29.2.16 Working with archived mail

Archived mail files can also be copied back to the original mail server from the WebClient. To do so, you use the `go_mail_xxx` command, where 'xxx' represents the file number. You can use `go_zip_xxx` to download the mail message as a zipped file.

29.2.17 Working with Office files

Archived Office files can also be restored from the WebClient. To do so, you use the `go_file_xxx` command, where 'xxx' represents the file number. You can use `go_zip_xxx` to download the Office file as a zipped file.

29.2.18 Determining the WebClient version

Each page that is delivered by the WebClient contains a hidden 'avversion' form variable that contains the current date of the release (e.g. 20080222). This allows the calling application to test whether a sufficiently updated version is available.

29.2.19 Automatic logoff after login

Using the 'reconnect=1' option you can arrange for the connection to be closed after each API command. Without this option, the connection is left open until the logout process is completed.

29.2.20 Hide login information

With the option 'redirect=1', when logging in, a second access occurs immediately so that the login information is not displayed in the browser.

29.3 Examples of working with API Interface

Below we present some examples of what is possible using the API of the WebClient.

29.3.1 Automated Uploads with WinUpload

For uploading a file the following options are available:

1. Upload of the revised file via WebClient
2. Upload via FTP folder (/office/dbname)
3. Upload via API interface

This last option is very powerful, but requires a small script (as shown in this example with our main language, Perl).

```
#!/usr/bin/perl
my $fout = shift; # get here the filename
use strict;
use LWP::UserAgent; # we work with UserAgent (our batch web browser)
use HTTP::Request::Common qw(POST); # the post method must be imported
my $server = "http://192.168.0.220/perl/avclient/index.pl";
my $www = LWP::UserAgent->new; # new www session
my $res = $www->request(POST "$server",
    Content_Type => 'form-data', # multipart/form-data
    Content => [ # structure for our file
        MAX_FILE_SIZE => 134217728, # max. size (WebClient won't accept more)
        upload => [ $fout, $fout ], # file to upload, file name to use
        go => 'go_action', # we need to call the go_action command
        action => 'upload', # inside of go_action we need to use upload
        uploaddef => 0, # scan def
        host => 'localhost', # connection information
        db => 'archivista',
        uid => 'Admin',
        pwd => 'archivista',
        frm_Laufnummer => 1,
    ]
);
```

29.3.1.1 Automation in Windows

These lines of code upload a file via our API to the WebClient. They can just as easily be entered in any other programming language. And yet, most of our customers do not want to deal with

such 'trifles' and therefore the demand for an automated upload directly from a Windows directory was high.

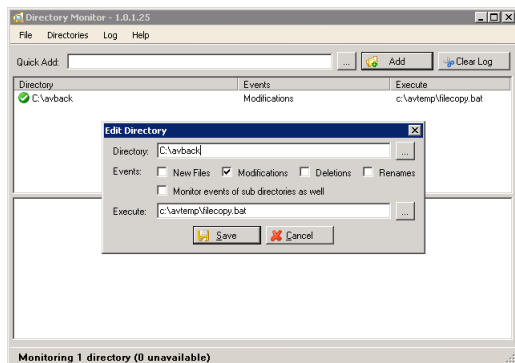
As a preliminary remark it should be noted that common Office packages can easily save to an FTP folder. There are also programs which can use any FTP folder as a drive path. Another option may be to monitor a directory and start an upload program when a modified file arrives.

Here is a small overview of such programs; all are available for free download on the Internet, e.g. at www.technixupdate.com:

- MoniDIR 2000
- Bee Watcher
- Folder spy
- FileGuard
- FolderMon
- SilentNight Inspector
- Directory Monitor

29.3.1.2 Directory Monitor monitors directories

We want to mention Directory Monitor as an example. It is important for this that you activate 'Modifications' when downloading.



Under 'Execute' you can then store an upload script in order to perform the upload. In this specific case we have used the following script, filecopy.bat:

```
cd c:\avtemp
@echo off
echo user ftp> ftpcmd1.dat
echo ftp>> ftpcmd1.dat
echo cd office>> ftpcmd1.dat
echo cd archivista>> ftpcmd1.dat
echo bin>> ftpcmd1.dat
echo put %1>> ftpcmd1.dat
echo quit>> ftpcmd1.dat
PING -n 3 127.0.0.1>nul
ftp -n -s:ftpcmd1.dat 192.168.0.220
PING -n 1 127.0.0.1>nul
del ftpcmd1.dat
del %1
```

DriveMonitor makes sure that after editing a file, the upload is done automatically. This method generally works fine.

However, there are two things to remember. Firstly, DriveMonitor requires a more recent Dot.Net environment. The application itself is reassuringly small, but the Dot.Net environment is frighteningly large (and does nothing without a reboot).

Secondly, DriveMonitor is a freeware application. This means that you may use the application free of charge, but do not receive the source and cannot, for example, judge how stable or good DriveMonitor really is.

29.3.1.3 Solution with WinUpload or WinCopy

The solution described in the previous subsection has another disadvantage. During the FTP upload a text file is created to perform the upload via the internal Windows FTP client (ftp.exe). In this, user names and passwords are stored in plain text (both in the batch file, and in the control file). In addition, filecopy.bat (our upload program) must be adjusted by hand.

For this reason we have decided to create two new utilities. These are winupload.pl and wincopy.pl. These programs are available as of 2010/IV Release in the /home/cvs/archivista/jobs folder. The two Perl programs are platform-independent. On Linux, Perl is probably already available, but not on Windows. For this reason, we have compiled the programs, i.e. you can use the programs as EXE files. The zipped versions can be found here:

www.archivista.ch/de/media/version/winupload.zip

and

www.archivista.ch/de/media/version/wincopy.zip

Using the first program (WinUpload), files can be uploaded easily.

An entry with no parameters prints a short help text:

```
C:\avupload\winupload.exe v1.0 (c) 2010 by Archivista GmbH, upload files to WebClient
This program has three modes, a) single, b) config and c) upload
a) single mode: C:\avupload\winupload.exe host db user password def https file [meta]
b) config mode: C:\avupload\winupload.exe host db user password [def=0..x https=0/1]
c) upload mode: C:\avupload\winupload.exe file [fieldname1=value1;fieldname2=value2..]
```

From this we can see that we can access WinUpload in three variants.

- In Single mode we enter all the parameters and the upload of the file occurs directly (without configuration).

- In the Config mode we can enter the upload parameters so that we can then only specify the file name.
- In Upload mode, finally, we only enter the file and any meta-information in order to upload a file. The connection options are read from the file 'winupload.dat'.

The file 'winupload.dat' is not available in plain text, but it must be stated that if you need high quality encryption, the present 1.0 version of WinUpload is not sufficient.

The following are the three modes, each with an example.

- Single-Modus: winupload 192.168.0.220 archivista Admin archivista 0 0 c: \14282d.pdf
- Config-Modus: winupload 192.168.0.220 archivista Admin archivista 0 0
- Upload-Modus: winupload 192.168.0.220 c: \14282d.pdf
Text = Hallo

Before we can upload a file, we need to monitor a directory. This is what the second program (wincopy) is for. You can start the program without parameters and then receive a list of options.

```
C:\avupload\wincopy.exe v1.0 (c) 2010 by Archivista GmbH, watch a dir and call upload prg
C:\avupload\wincopy.exe dirin dirtemp seconds (to wait between checks for uploads) [upload]
=> [upload] optional, without it call winupload.pl or winupload.exe
=> you can stop the program if you create a file wincopy.stp
```

➡ **Important! We strongly recommend that you only start the program within an empty application folder. If 'wincopy.exe' opens with a valid path, then all files will immediately be moved to a temporary directory and an attempt will then be made to upload the files.**

But back to the parameters. The monitored directory must come first. This must be followed by a temporary directory in order to

check whether a file to be uploaded is still 'locked' by an application. In this case the upload cannot take place.

Third, a pause time can be entered in seconds (3 - 60). During this time, the directory indicated previously will be examined. Finally, here is a complete example:

1. Create a folder 'c:\avupload'
2. Copy the 'winupload.exe' and 'wincopy.exe' files to this directory
3. Start wincopy.exe 192.168.0.100 archivista Admin 0 0 (or other login options)
4. Check the 'c:\avupload' directory to see whether the file 'winupload.dat' exists
5. Verify that the 'c:\temp' directory exists (otherwise create it!)
6. Start 'wincopy.exe. '(with the dot at the end you can start your current directory)
7. Save an Office file in the 'c:\avupload' directory, and the file is then uploaded immediately.

➡ **Note:** Item 7 only works if the file is no longer locked by the processing program. You will generally need to 'Save' or even 'Close' the file.

Of course you will want to open the 'wincopy.exe' utility when booting or via a service. And of course you may want to adapt the Perl programs 'wincopy.pl' and 'winupload.pl' to fit your personal needs, i.e. if you want to place the source code under GPL license once again.

29.3.2 Automated indexing with PDF-Creator

This feature requires the very latest ArchivistaBox (20100308 Release) and the 'API' module, so that we can provide support for the module. In addition, the 'PDF Creator' virtual printer is required.

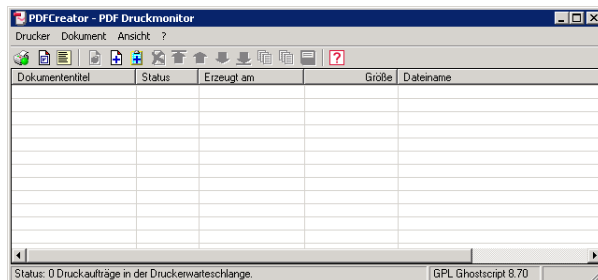
The program is open source, and you can obtain the current 0.9.9 version at sourceforge.net. PDF-Creator is a virtual Windows printer which creates PDF files. PDF Creator can also call up other programs in the background after creating a PDF.

You can use this functionality in order to create a PDF file using PDF-creator, upload it to the ArchivistaBox and then access the WebClient in order to directly access the previously uploaded file or immediately create an index.

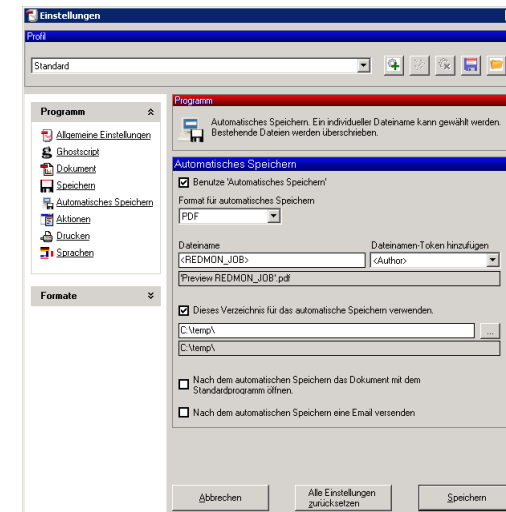
This allows you to index the printed files directly after printing, without having to first search for the file.



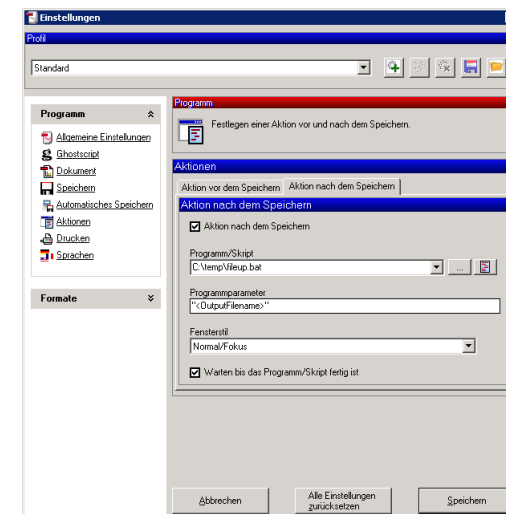
The following is a short guide to the use of the program. First, install PDF Creator. After installation, you must restart PDF-Creator. You will see the following form:



Select 'Settings' under the 'Printer' menu item. You will see a form. Within this you will see 'Automatic saving'. Select this option to arrive at the following form:



Please note that we automatically save the PDF files created to c:\temp. For this we select 'Actions'.



We specify here that in the `c:\temp` directory the batch file `fileup.bat` is to be run. Here, we want to publish the file as a text file. You must adapt it to your environment if necessary.

```
@echo off
echo user ftp> ftpcmd.dat
echo ftp>> ftpcmd.dat
echo cd office>> ftpcmd.dat
echo cd archivista>> ftpcmd.dat
echo bin>> ftpcmd.dat
echo put %1>> ftpcmd.dat
echo quit>> ftpcmd.dat
ftp -n -s:ftpcmd.dat 192.168.31.112
del ftpcmd.dat
del %1
cd C:\Programme\Mozilla Firefox
firefox.exe "http://192.168.31.112/perl/avclient/index.pl?host=localhost&
db=archivista&uid=Admin&pwd=archivista&go_queryfile&fld_Dateiname=%1"
```

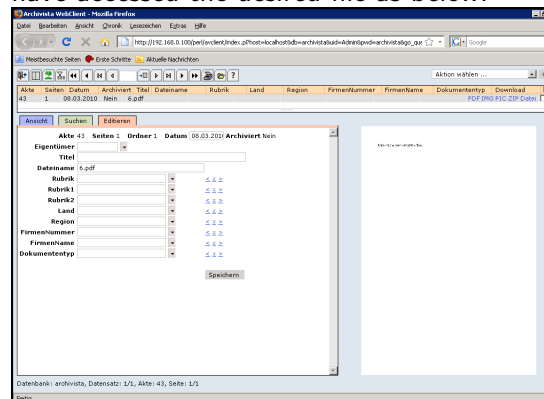
Copy the above text into notepad or another text editor. Save this file as a batch file (ending with 'bat') into the `c:\temp` directory. It is best to then click on the 'fileup.bat' file in Explorer, and the batch file should have started (brief appearance of a black window).

What does this file do now? First, the `ftpcmd.bat` file is created. It is necessary to stipulate in it that the file obtained is to be uploaded via FTP to the ArchivistaBox. Thereafter, the 'ftp' program must be called up and the current `ftpcmd.dat` file is transferred. Please enter the correct IP address of the ArchivistaBox into the FTP server, and ensure that the FTP server is switched on.

After the upload the current `ftpcmd.dat` file is deleted, as is the resulting PDF file. The script will now switch to the Firefox folder (enter other browsers here if necessary) and start the WebClient, whereby the desired user must be registered. The last piece of information calls up the previously uploaded file within Archivista. For this to work, you must apply the filename to a field in WebAdmin. In our example we have used the 'FileName' field. How to assign a name to a field in WebAdmin is explained in 12.

So, now you can print from any application using the 'PDF Creator' printer. The file is then automatically loaded onto the

ArchivistaBox and the WebClient is launched. You should then have accessed the desired file as below.



You can now index the file. Then click 'Save' and the file is indexed.

30 Database structures

30.1 The most important Archivista tables

In the following we would like to introduce you to the most important tables of an Archivista database. Please note that table and field names are in German as yet. For a conversion table of field names please see chapter 30.5.

- **Table 'archiv':** Each record corresponds to an Archivista document. The field 'Laufnummer' (English: 'counter') administrates the document number. The field 'Gesperrt' (English: 'locked') is essential in that each time a user tries to change a document the system checks this field to find out whether or not somebody else is already working on the same document. If the field 'Gesperrt' is empty, the user's ID is entered in the field and he is free to make changes. After he has finished the content of the field 'Gesperrt' is erased again.
- **Table 'archivseiten':** Each record corresponds to a page in a document. The necessary link between documents and pages is saved in the field 'Seite' (English: 'page'): the last three digits denominate the pages of a document while the preceding digits stand for the document number. Let us make an example: '2412004' can be parsed into document number '2412' and page number '4.' It is important to know that there is no relational mapping between the two tables as yet. That is, an application must take care, for example, that all pages in table 'archivseiten' are deleted, the page counter in table 'archiv' (field 'Seiten') is updated and the document itself is deleted (if there are no longer any records in table 'archivseiten').

- **Table 'parameter':** In this table you find all settings which play a role in any Archivista database. Most relevant are certainly the entries regarding the archive directories. These entries begin with 'Pfad ...' (English: path).
- **Table 'user':** In this table you find the user specific settings. For more information see 30.3.

30.2 Archiving concept

30.2.1 Tables and Fields

30.2.1.1 Table 'archive'

In table 'archive' the individual documents are stored. The keywords, the number of pages and the folder number are to be found here, for example, but not the image files or texts of the individual pages.

```
mysql>
mysql>
mysql> select Akte,Seiten,ArchivArt,Erfasst,Archiviert,BildInput,BildIntern,QuelleIntern,BildInputExt,BildExt,QuelleExt from archiv order by Akte desc limit 3;
```

Akte	Seiten	ArchivArt	Erfasst	Archiviert	BildInput	BildIntern	QuelleIntern	BildInputExt	BildExt	QuelleExt
77	230	1	1	0	0	0	0			
76	16	1	1	0	0	0	0			
75	1	1	1	0	0	0	0			

```
3 rows in set (0,00 sec)

mysql>
```

Let us take a look at the relevant fields of table 'archive':

- Dokument: unique number for document (1-x)
- Seiten: Number of pages per document (0=none,1-640 pages)
- ArchivArt: 0=BMP/ZIP (obsolete), 1=TIF, 2=PNG, 3=JPG
- Erfasst: 0=no pages added, 1=document contains pages

- Archiviert: 0=document is not archived, 1=document is archived
- BildInput: 0=image in 'input' folder, 1=image in table 'archivbilder'
- BildIntern 0=no web preview images, 1=web preview images
- QuelleIntern: 0=no source file, 1=source file
- BildInputExt: file extension in capital letters (TIF,PNG,JPG)
- BildAExt: file extension for web preview images (TIF,PNG,JPG)
- QuelleExt: file extension for source format archiving (e.g. DOC,PDF)

30.2.1.2 Table 'archivseiten'

Table 'archivseiten' contains the information regarding the text recognition of a page, i.e. we find the extracted text as well as the OCR definition belonging to the page. Note that the archiving process leaves table 'archivseiten' untouched.

```
mysql> describe archivseiten;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Seite | int(11) | | PRI | 0 | |
| Ausschiessen | tinyint(4) | | MUL | 0 | |
| Erfasst | tinyint(4) | | MUL | 0 | |
| Schlüssel | varchar(56) | | | | |
| Text | mediumtext | YES | MUL | NULL | |
| Zipped | tinyint(4) | | MUL | 0 | |
| Indexiert | tinyint(4) | | MUL | 0 | |
| OCR | int(11) | | | 0 | |
| Notes | blob | YES | | NULL | |
| ScreenQuality | int(11) | | | 0 | |
+-----+-----+-----+-----+-----+-----+
10 rows in set (0,06 sec)

mysql>
```

The following fields are relevant for us:

- Seite: reference to document and page of table 'archive' (document*1000+page)
- Ausschiessen: do not treat page with OCR
- Erfasst: page already treated with OCR
- Text: memo field for various information
- Indexiert: shows whether an index has been done for this document (obsolete)
- OCR: definition (0-x) making the desired language strings available
- ScreenQuality: reduction factor for screen copy: 0-50

30.2.1.3 Table 'archivbilder'

Table 'archivbilder' houses the individual image and source files.

```
mysql> describe archivbilder;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| Seite | int(11) | | PRI | 0 | |
| Bild | mediumblob | YES | | NULL | |
| BildA | mediumblob | YES | | NULL | |
| BildInput | longblob | YES | | NULL | |
| Quelle | longblob | YES | | NULL | |
| BildX | int(11) | YES | | NULL | |
| BildY | int(11) | YES | | NULL | |
| BildRX | int(11) | YES | | NULL | |
| BildRY | int(11) | YES | | NULL | |
| DatumA | datetime | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
10 rows in set (0,07 sec)

mysql>
mysql>
```

Let us take a look at the relevant fields of table 'archivbilder'

- Seite: reference to document and page of table 'archive' (document*1000+page)
- Bild: first image file existing in reduced web quality
- BildA: currently not used
- BildInput: original image file, i.e. page as captured

- Quelle: source file (either from first page or each page individually)
- BildX, BildY: pixel data of file in field 'Bild' (optional)
- BildAX, BildAY: pixel data of file in field 'BildInput' (optional)

30.2.2 Archiving process

The objective of the archiving process is to save the added documents in a way that they cannot be altered anymore and that they remain readable from a long-term perspective.

That image data become unalterable is effected by the field 'Archiviert' in table 'archive' showing '1': any manipulation of the image in an Archivista application is no longer accepted.

Long-term readability of the image data is achieved by the fact that in any case they are saved in a certain folder structure and that the images thus created, or the corresponding folders, respectively, are written to non-rewritable media (CDR/DVDR).

30.2.2.1 Purpose of folder number

The field 'Folder' of table 'archive' serves to store the information in which folder a specific file (image file) can be found. The folder number is managed by the system. The last, i.e. most recent folder number is managed in table 'parameter'. There the field 'Name' has an entry 'ArchivOrdner'. The number found there is the currently used folder number.

In principle each document is allocated the current folder number when the document is being created. The folder number refers to the corresponding archiving folder. Example: if document 84 shows folder number 10, it will be archived in folder 'ARCH0010' if the previous documents have found room too.

Important for the archiving process are the following additional entries in table 'parameter':

- ArchivAutomatisch: 0=manual (obsolete), 1=automated archiving process
- ArchivMByte: 100-10000 (value=number of MByte per archive folder)
- ArchivDateien: 100-10000 (value=number of files per folder)

In 'ArchivMByte' one can determine how many MByte are to be saved in a folder (e.g. 600 MByte), in 'ArchivDateien' one can specify the maximum number of files (e.g. 2000 files) that are allowed in a folder.

In the following we give two examples.

First example: For the moment we do not know if document 84 can still be archived in folder 10 (name: ARCH0010). Currently, folder '10' is made up of 25 MBytes and 1789 files. The field 'ArchivDateien' shows 1800 files, our document 84 possesses 14 pages. Together with document 84 folder 'ARCH0010' would contain 1803 files. That is the reason why a new archiving folder with the name 'ARCH0011' is created and the current folder number goes up to 11.

Second example: Up to and including document 83 there are 1500 files in folder 'ARCH0010'. The folder consists of 589 MBytes, document 84 contains 220 files and is made up of 44 MBytes. Also in this case document 84 would be finally saved in archive folder 'ARCH0011'.

30.2.2.2 Calculation of file name

If a page is saved in the database itself the image file is not named. The file name is calculated on the basis of the field 'Document' and the corresponding page. It works like this:

The image file gets an 'A' in first position. The second to the sixth position are reserved for the document number, the seventh and eighth position gives the current page.

The following example will make this clear.

Document number 84: modulo 26 = 6 (G) remainder, 3 (D) results in: 000DG

Page number 12: modulo 26 = 12 (M) remainder, 0 results in 0M

Thus, the full file name (without extension) looks like A000DG0M.XXX. The extension of the file name is calculated on the basis of field ArchivArt:

- 0 = BMP/ZIP
- 1 = TIF
- 2 = PNG
- 3 = JPG

If field 'ArchivArt' displays the value '3' all this results in the file name A000DG0M.JPG.

30.2.3 Differences in the archiving process

30.2.3.1 Working with internal images (ArchivistaBox)

- The image files are immediately stored in table 'archivbilder' and field 'BildInput'.
- The source files are immediately stored in table 'archivbilder' and field 'Quelle'.
- The field 'BildInput' in table 'archive' shows '1'.
- During the archiving process the files are always moved to folder 'output'. The image files can remain in the database but they need not remain there.

30.2.3.2 Working without internal images (Archivista 5.x)

- When pages are added the image files are saved in folder 'input'
- Newly added source files are also filed in folder 'input'
- The field 'BildInput' of table 'archive' shows '0'.
- During the archiving process the files are moved from folder 'input' to folder 'output'.

30.2.4 Archivista folders

When Archivista archives documents it counts the number of pages present in the current directory. As soon as a certain total of pages is reached a new directory is automatically created and the surplus pages are copied there. The reason for this is that an operating system can only handle a certain number of files per directory efficiently. A new directory is created, too, when the size of the currently used directory exceeds the predetermined size of the external device. If no new directory were created in such a case, it could not be transferred later.

As mentioned previously the task of handling pages and preparing them for transferral to external media is taken over by Archivista. The only thing you should keep an eye on is the folder number. If after archiving the number in the field 'folder' is increased, a new folder has been created and the old one (only that one) may be transferred.

An example: The fields 'folder' of documents 1 to 34 show a 1 before archiving. After the archiving process the documents 29 to 34 have a 2 in their respective field 'folder'. Folder 1 is definitively full and may be transferred.

If there is an output directory under 'C:\Program Files\av416e\archive\output\' you will find the two subdi-

rectories 'arch0001' and 'arch0002' in them after the archiving process has been run. The first subdirectory, 'arch0001', can be transferred to an external storage device because it is complete ('arch0002' is incomplete and should be left alone).

➡ If you work with standard settings the folders are made to fit on CDR. You need not make adjustments to the parameters for archiving.

A further piece of advice: To be on the safe side with regard to data security you should make two copies of the complete folder that you want to transfer.

30.3 Clean login to Archivista databases

At the beginning is a clean login to the system. Basically you need the required MySQL rights to access an Archivista database. To allocate these rights you either create a user on the MySQL console or you use an ordinary Archivista user account.

The first alternative is probably easier to implement while the second gives you more flexibility. With the second alternative you can fall back on the existing user specific information, something you cannot do when using the first. The user specific information is to be found in table 'user.' This is the same for all Archivista databases.

When login is attempted through an Archivista client the system checks whether the user possesses access rights (reading rights as a minimum) for a particular Archivista database. The MySQL server consults its internal administration database 'mysql' and takes a look at the entries in table 'tables_priv.' If user ID and password are correct, the user may access the database. In a second step the system checks the user specific information in table 'user' of our Archivista database. Is there an entry which has the same information in fields 'user' and 'host' as the internal MySQL table

'tables_priv'? If this is the case, too, login may definitively take place.

30.4 File name of an external saved page

After we now know how to access an Archivista database we would like to show an exemplary case of how to locate bitmap pages with an SQL query and a few commands in Perl.

```
sub GetDateiName {
    my $modus = shift; # A or B copy
    my $akte = shift; # document number
    my $seite = shift; # page in question
    return $modus . FileName($akte,5,) . FileName($seite,2);
}

sub FileName {
    my $eingabe = shift;
    my $anzahl = shift;
    my ($res,$res1,$sout);
    while ($eingabe != 0) {
        $res = $eingabe % 26;
        $res1 = $res + 65;
        $sout = chr($res1).$sout;
        $eingabe = $eingabe - $res;
        $eingabe = $eingabe / 26;
    }
    $sout = "0" x ($anzahl - length($sout)) . $sout;
    return $sout;
}
```

The code above shows how the file name is calculated from the document number and the page number in that document.

To calculate the complete path of a page we in addition need the type of the document and the folder where it lies. The type of the document is to be found in field 'ArchivArt': 0=BMP/ZIP,1=TIF,2=PNG and 3=JPG. The folder number is to be found in field 'Ordner.'

One other thing must be paid attention to: has a document been archived or not? The field 'Archiviert' in table 'Archive' tells us. If yes, the page will no longer be found in the input folder but it will reside in the output folder. The folder of a document that has not yet been archived can be located by the following procedure:

a) it must be in the input directory b) its number is the result of taking the remainder of modulo 100 of a document number (e.g. the remainder of 1725 would be 25).

Two more examples:

Page in input folder: Document 1, page 1 with A copy and value '1' in field 'ArchivArt' you find in '..\input\in001\A0000B0B.TIF'.

Page in output folder: Document 2, page 3 with B copy and format 'JPEG' has already been archived and lies in folder '1' (field 'Ordner,' English: 'folder'). This page is to be found in '..\output\ARCH0001\B0000C0D.JPG'.

➡ The Archivista PublishingEdition offers more clues as to how to communicate with an Archivista database. It gives you the complete source code (including correct login) for communicating successfully with Archivista databases.

30.5 Field names

Due to the fact that the original application was written in German the database coded field names are in German. To ensure compatibility between the different language versions the field names are kept in German. For modifications in the 'Definitions for selection' and the 'Print reports' you have to work with the original field names. In the following you find a conversion table:

German=English

Titel=Title

Datum=Date

Akte=Document

Seiten=Pages

Stichwörter=KeyWords

Personen=Persons

Notiz=Note

ErfasstDatum=EntryDate

Ordner=Folder

Farbe=color

EDVName=FileName

Erfasst=Added

Archiviert=Archived

Eigentümer=Owner

Markiert=Marked

Laufnummer=Counter

NotizRTF=NoteRTF

Verbindungen=Hyperlinks

Gesperrt=Locked

Anrede=Address

Vorname=FirstName

Nachname=LastName

Zusatzzeile=OptionalLine

Strasse=Street

Landcode=Countrycode

PLZ=ZIP

Ort=Town

Land=Country

Telefon=Phone

Geschäft=Business

Telefax=Fax

Zusatz=Phone2

Geburtsdatum=BirthDate

Temporär=Temporary

Briefanrede=Salutation

Bemerkungen=Remarks

Aufnahme=EntryDate

BemerkungenRTF=RemarksRTF

DatumVon=DateFrom

PendentAb=PendingFrom

Erledigt=Done

Betrifft=Subject

NotizRTF=NoteRTF

Untertitel=Subtitle

Autoren=Authors

Verlag=Publisher

Auflage=Edition

Sprache=Language

AnzahlSeiten=PagesOfBook

Ausgabejahr=YearOfPublication

ISBNNummer=ISBNNumber

DatumKauf=Purchased

Preis=Price

Gebiet=Region

Note: The end user does not see any German field names, everything will be translated to English field names by the program.

Part XI

FAQs

31 ArchivistaBox

31.1 Is there a demo version of Archivista?

We shall be pleased to make demo versions available – please contact us.

It is considerably simpler when you use the free-of-cost options. In the case of ArchivistaDMS this is ArchivistaCloud and in the case of ArchivistaVM it is the ArchivistaBox Mini. This choice of products can be used for free anytime (after registering with our webshop).

31.2 Is ArchivistaBox free?

Short answer: Yes. Long Answer: No software is free in the broader sense. The development of any software costs time and money. We specifically decided for a GPL license in 2005 because we believe that open sources (programs) are beneficial for both our customers (trust) and us (our work goes into the solution, not the license management). Furthermore, interested parties can participate in the project. This does not, however, mean that everything in the ArchivistaBox falls under the GPL license. For example, the manuals are not subject to the GPL license and are therefore subject to a copyright. Put briefly, if it says Archivista on it, there must be an ArchivistaBox inside. Please observe the fair use principle according to 1.5.1.

31.3 ArchivistaBox does not start up properly

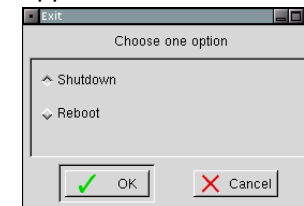
Should the ArchivistaBox freeze while booting up, we recommend that you install an update or boot up the old (second) installation.

31.4 Shutdown of ArchivistaBox

The ArchivistaBox is an archiving server designed for 7x24 hours' running. Nevertheless it may make sense under certain circumstances to turn it off.

Click the X in the upper right hand corner of the screen to change to the level of the ArchivistaBox settings. You will see a grey screen. Right-click your mouse and choose 'Exit' from the context menu that appears.

You are asked for the root password. Then, a new dialog box appears.



Click 'Shutdown' or 'Reboot' depending on your wishes.

The archiving server will shut down.

➡ Please note that there is no need to press any buttons for turning the ArchivistaBox off.

31.5 Installation/Update from CD

Installation is no longer necessary with ArchivistaBox, the system is started up from CD/USB stick. Normally, the corresponding ISO file is saved directly on ArchivistaBox (the internal hard disk). To start-up a new release, you usually make use of the online update possibility. Cf. 5.3

Should the online update fail (this is, for example, the case when ArchivistaBox has no access to the Internet), each ArchivistaBox

can be started from CD/USB stick. Instructions of how to obtain your personal ISO file should have reached you by means of a separate document. Should no external CD drive be available to start ArchivistaBox from, take recourse to our website www.archivista.ch where the creation of a USB stick is extensively described.

31.6 Installing ArchivistaBox on a Virtual Machine

You can operate ArchivistaBox in virtualised form anytime. Currently, the server products from VMWare, KVM (e.g. ArchivistaVM), Hyper-V and XEN are supported. How an operating system is installed (in our case Linux 64bit) can be gathered from the manual that comes with it.

31.7 Saving your data

➡ Please note the difference between long term and short term data saving. The ArchivistaBox concept supports the following levels of data security.

- Mirroring to second ArchivistaBox to save changes made during the current day (only ArchivistaBoxes 'Titlis' and 'Eiger').
- Daily backup by means of a) USB stick or b) tape
- Storing of data on non-rewritable media like CDs or DVDs for longterm data storage.

Mirroring to a second ArchivistaBox is described in 25.13. The backup is described in section 25.8.

We recommend (and many countries require it) that you store your data on non-rewritable media for the purpose of safe **long-term keeping**. The procedure is described in more depth in 25.6.3.

31.8 Data exchange and backup by means of USB stick

31.8.1 General

We recommend that you buy USB sticks. USB sticks play an important role in the data exchange between ArchivistaBox and other digital devices and they are very easy to use. Depending on the action you want to take you create a folder on the formatted stick (e.g. 'exchange') and you connect the stick with your ArchivistaBox. On the basis of the folder name ArchivistaBox will carry out the action you chose. In principle, an empty folder on a USB stick leads to an export from your ArchivistaBox while a folder containing data incites ArchivistaBox to import data. Currently, there are the following types of action:

pdf	PDF export of specific documents
exchange	Import/export of documents
crypt	Import/export (encrypted)
transfer	Import/export between box and RichClient
backup	Backup (for smaller archives)
cust	Cust folder (user specific scripts)
config	All parameters from an ArchivistaBox
'other name'	Import of images to Archivista

In the following the procedure is described for the various ways of exporting and saving data. The procedure for the saving of data as PDF files is described in detail. The rest works as described for PDF files.

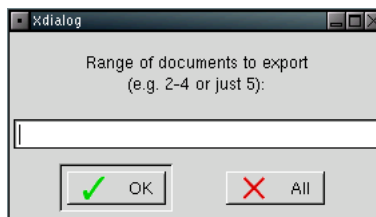
31.8.2 PDF

As explained above the ArchivistaBox recognizes what action it is to take by the properties of the USB stick. For a PDF export to take place the USB stick must be prepared accordingly.

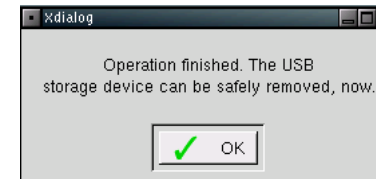
Go through the following steps:

- Attach the USB stick to your Windows PC
- It appears as 'Removable medium' on your PC
- Right click your mouse: format as Fat32
- Right click your mouse: create a directory / folder named 'pdf' on your USB stick (lower case)
- Right click your mouse: 'Eject'
- Remove the USB stick and insert it in the USB interface at the front of your ArchivistaBox

After entering the Archivista Admin user password you see the following dialog box:



Enter the range. (The figures 'from-to' refer to the document numbers.) Now the ArchivistaBox copies the documents onto the USB stick. When the process is finished the following message will come up:



31.8.3 Ordinary Archivista export for later re-import to ArchivistaBoxes

Proceed in the same way as described above in 'PDF export' but instead of creating a directory named 'pdf' create one called 'exchange'.

After you have entered the Archivista Admin user password the copying process is carried out. When the process is finished you will see a message. The export has succeeded.

31.8.4 Archivista export for later re-import to RichClient

Proceed in the same way as described above in 'PDF export' but instead of creating a directory named 'pdf' create one called 'transfer'.

After you have entered the Archivista Admin user password the copying process is carried out. When the process is finished you will see a message. The export has succeeded.

31.8.5 Backup

Proceed in the same way as described above in 'PDF export' but instead of creating a directory named 'pdf' create one called 'backup'.

After you have entered the Archivista Admin user password the copying process is carried out. When the process is finished you will see a message. The backup export has succeeded.

31.8.6 Customer specific adjustments (cust)

Each ArchivistaBox has the same structure. Sometimes an ArchivistaBox is supposed to fulfill a very specific function for a customer - a function which the customer does not wish to publish as OpenSource license or one which we cannot incorporate into the official ArchivistaBox (e.g. for reasons of space).

Further cases are: form recognition and specific programs for the processing of print data (COLD).

In all these cases we use the `cust` folder together with a USB stick. If the folder is empty, all available scripts/data are exported. If the folder is not empty, the data from the USB stick are copied to the ArchivistaBox.

31.8.7 Image import from digital cameras (other folders)

If the stick does not show a specific folder, ArchivistaBox searches for JPG images and, if available, offers to import these. This function is described in detail in 32.7.

32 Integration questions

32.1 Is there a mail archiving function?

Yes. An IMAP-compatible mail archiving function that can monitor and archive any folders is available for the ArchivistaBox. Archived mail can also be checked out again. See 25.12.1 for more information.

32.2 Can I archive Office files (incl. check out)?

Yes. As of release 2009/I, the ArchivistaBox can process more than 200 file formats without using external plug-ins. See 25.11.1.1 or 9.1 for more information.

32.3 Which possibilities are there with regard to scanning to ArchivistaBox?

You can choose among a direct connection to ArchivistaBox, ArchivistaBox scan stations and network scanners. Scanning with a network enabled device is outlined in 25.11. It is considerably simpler to directly attach a scanner to ArchivistaBox or to work with scan stations.

If a scanner is connected to an ArchivistaBox (e.g. ArchivistaBox Dolder), scanning will work without any further effort. The only prerequisite is that you have purchased the ArchivistaBox scanner drivers.

Also the scan stations need not be configured. Nonetheless, a user representing the scan station must be created in the appropriate database on the main box. To to **WebAdmin**, **User** and select

New on the upper left of the table. Enter the IP address of your scan box in field 'Host'. User and password are utilised in the same way as with other scan boxes and databases. Relevant information on this is given to you when the ArchivistaBox(es) is(are) delivered.

☞: In case you work with a keypad, all databases must be registered with the scan key (for more information see 25.14.1.8). This should be correctly preset, though, when the box is delivered.

32.4 Can I scan books?

A book can be scanned by means of the usual scanning command. Special attention should be paid, however, to a number of points. To create a scan definition for a book go to WebAdmin and menu **Scanning**. Click **New** to make a scan definition that fits the specific needs of the book to be scanned. For a detailed description of all scan definitions please read Chapter 15.

When scanning books it is fundamental to give the exact format. The scale of measurement is in millimeters. Inexact specifications may lead to a considerable impairment of scanning quality. Enter the measures in the following fields:

Scan width (mm)	<input type="text" value="210"/>
Scan height (mm)	<input type="text" value="297"/>

As most books are printed in black and white, the specifications **Grayscales** and 200 dpi are sufficient. If the books are printed in color, these values must be adjusted accordingly.

Color information	<input type="text" value="Grayscales (8 bit)"/>
Resolution (dpi)	<input type="text" value="200"/>

To achieve a really high scanning quality the option **Black/white optimization** should be ticked.

Maybe you are asking yourself why you do not scan in black and white. The answer is that black and white optimization does not work with a black and white scan.

Depending on how you scan your book it must be rotated in order that the scan is in an upright position later. For this purpose use the field **Rotation**. The scan is turned automatically in accordance with your settings.

Rotation

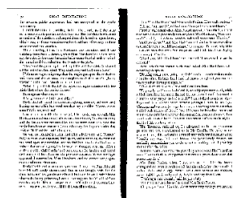
The option '**Several pages**' relieves you from the necessity to start the scanning command again and again. Under **Autopilot** one can specify how many scans should be made.

Several pages
Autopilot: number of pages

In the example above a '4' has been entered in the field **Autopilot: number of pages**. The scanner makes 4 scans. Between the scans the scanner takes breaks. By means of **Autopilot: break in seconds** it can be specified how long that break ought to be. The illustration below shows an example where the scanner is to wait 2 seconds in between scans.

Autopilot: break in seconds

When scanning a book you mostly scan two pages at once.



The function **Split page in middle** gives you the possibility to split the pages in two after scanning.



Since books are in most cases made up of text one should not forget **text recognition (OCR)**. Under 'Desired OCR definition' you can choose between Multi, German, English and German, and others.

32.5 Classifying documents (ERP interface)

You use the Archivista printer to print documents in the archive. You then want to index these documents automatically and you use Archivista to search for a solution. This solution is called 'coldplus.pl'.

/home/data/archivista/cust/cold/coldplus.pl

If this program exists in the ArchivistaBox, it is run for each print job. For example:

```
#!/usr/bin/perl

my $db = shift; # database name (you can change it)
my $title = shift; # title of document (goes to Title)
my $pdffile = shift; # pdf file name (for own work)
my $psfile = shift; # ps file name (for own work)
my $for = shift; # author (goes to Owner)

if (length($title)>120) {
    $title = substr($title,0,119);
}
$title =~ s/Microsoft Word - //g;
```



```

$title =~ s/Microsoft Office //g;
$title =~ s// /g;
$title =~ s:// /g;
$title =~ s=/ /g;

# to return values, you need to print it
# at first position include db name, followed by a ;
# after it add the field=value, separated with a :
my $ret = "$db;Title=$title:owner=$for";
print $ret;

```

The program always receives the database name, the respective PDF file, the corresponding postscript file and the author (if the author is included in the PDF document). The program initially returns the database name as well as the desired fields.

➡: ftpplus.pl exists in addition to coldplus.pl. The difference here is that only the name of the document is received. You have to add the corresponding program to the following location in the ArchivistaBox:

```
/home/data/archivista/cust/ftp/ftplusplus.pl
```

32.6 Is there an API for the ArchivistaBox?

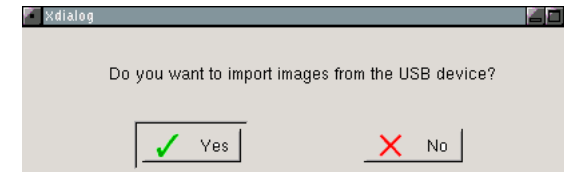
Yes. Use the external control for the Archivista WebClient. See 29 for more information.

32.7 Import of images

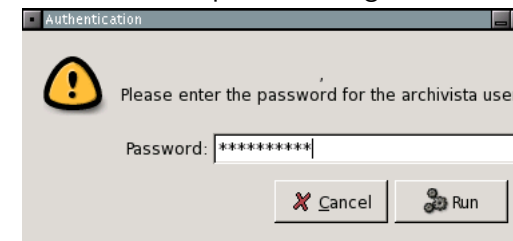
There is the possibility to store images from a digital camera directly on your ArchivistaBox. Attach your digital card or a USB flash SD card reader to the USB interface on the front of your ArchivistaBox.

32.7.1 Start import process

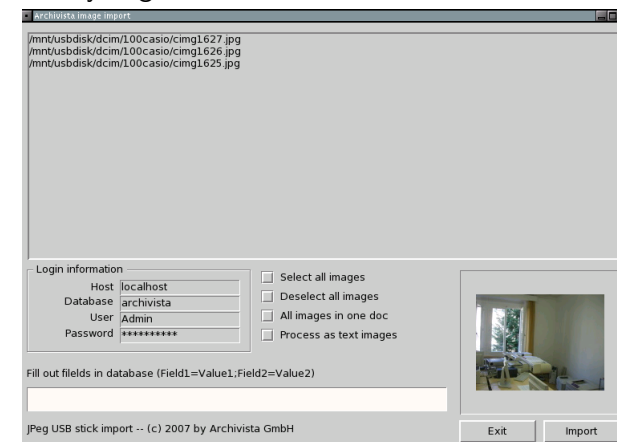
When the computer notices that you want to import image files you will see the following screen:



Click **Yes** to import the images. Then enter your password:



Now you get to the main menu:



You have the following choice:

- **Select all images**
- **Deselect all images** (= undo selection of all images)

- **all images in one doc** (= file all images in one document)
- **process as text images** (= import photographed text pages)

The field **Fill out fields in database** serves for the allocation of keywords to the individual photos. You may enter something like: Title=Office.

Fill out fields in database (Field1=Value1;Field2=Value2)

Title=Büro

JPEG USB stick import -- (c) 2007 by Archivista GmbH

Subsequently, the column 'Title' in the WebClient will show the name of the image: **Office**.

Akte	Seiten	Datum	Archiviert	Publizieren	Titel	Kontinent	Geschichte	Land	Region	Firmennummer	Firmenname	Dokumenttyp	PDF
172	1	16.07.2007	Nein		Büro								Download

Enter the name of the column in which you want to see the keyword. Enter the equals sign and then add the keyword you want. In the bottom left corner you see the following dialog box:

Login information

Host: localhost

Database: archivista

User: Admin

Password: *****

Normally this box shows the address of your ArchivistaBox. If you want to access a different ArchivistaBox you can change the data here.

To import the images click **Import** and, after that, **Exit** to leave the menu. A message appears telling you that you can remove the USB medium. Click **OK**.

32.7.2 Check the imported images in WebClient

The WebClient gives you the possibility to change to the so-called picture mode. Click the **tree symbol** in the top bar.



Your photos will appear as in a photo gallery, the evident advantage being that you can view all photos at once and can easily switch from one to the other. You can leave the picture mode by again clicking the **tree symbol**.

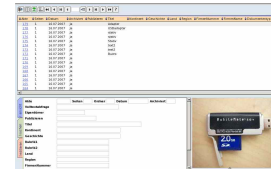
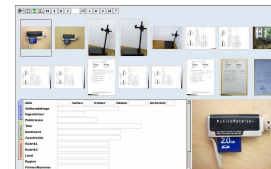


TABLE MODE



PICTURE MODE

What is more, the WebClient - both in picture mode and normal mode - lets you retrieve the **Exchangeable Image File Format Information**, i.e. the EXIF information. Simply click the blue **View** tab.

Dokumententyp

File name : /mnt/usbdisk/dcm/100casio/cimg1620.jpg

File size : 642970 bytes

File date : 2007-07-16 08:07:26

Camera make : CASIO COMPUTER CO.,LTD.

Camera model : EX-5500

Date/Time : 2007-07-16 08:07:25

Resolution : 1600 x 1200

Flash used : No

Focal length : 6.2mm (35mm equivalent: 38mm)

Exposure time : 0.0010 s (1/1000)

Aperture : f/4.3

Whitebalance : Auto

Metering Mode : matrix

Exposure : program (auto)

32.7.3 Photos with text

The import of photographed text documents works in analogy to the import of photos, however, there is one difference. In the main menu of the photo import feature click **Process as text images!** The photo is imported as usual and is archived by ArchivistaBox. Prerequisite for a good 'scan' of the text document is a good

quality, high-resolution camera. Many digital cameras possess a **text mode** specifically intended for the photography of texts: go to the menu 'settings' of your camera. Additionally, a camera support may be a definite advantage as pictures taken without support are easily blurred.



Such a camera support may be purchased from Archivista GmbH.

A camera of good quality (with, say, 10 megapixels, but even 5 megapixels may be sufficient) will return a photographed document of excellent quality. Let yourself be convinced by the following example:

```

-----
Désignation d'article          cc
Attention: prix et catalogues valables just'c
la fin de fevrier
-----
JS 069.03.001.0 400011          CH Pce.
Kränzle Nettoyeurs haute pression Flyer 2005

```

If the quality is sufficient (the letters must be closed and easily readable), the text recognition works also with a photographed document and the text of the document can be found behind the blue **View** tab in the WebClient.

```

Service de l'épa r a l : ion et; poe do ruoli. *Pel 052 3541444 Kax: 052 3543445
Metabo
COPE
Quincaillerie Riviera SA Andrey Philippe Avenue Nestlé 25 Case postale 1800 Vevey 1
Quincaillerie Riviera SA Andrey Philippe Avenue Nestlé 25 Case postale 1800 Vevey 1
Bulletin de livraison
Pour info : Date :
Claudio Togni 052/3543 16.01.06 Page
N/no de ode No de client Votre commande du
1290562-1174418 11.01.06 par représentant
Votre référence de commanc Mr. Vallotton
Désignation d'article
commandée
Attention: prix et catalogues valables just'qua la fin de février
livrée
Prix à la p:
10 AUS 069.03.001.0 400011 CH Pce.
Kränzle Nettoyeurs haute pression Flyer 2005

```

32.8 Transferring the RichClient to the ArchivistaBox

The easiest way to transfer an existing RichClient database to the ArchivistaBox is to carry out an export from the old Archivista database and then an import using a USB stick. You have to save the exported data in the 'transfer' folder. This is the only way to ensure that the USB stick is recognised correctly in the ArchivistaBox.

🔄: Please note that the field information is only transferred if the database in the ArchivistaBox has the same fields as those of the old RichClient database.

32.9 Importing data from other DMS software

The ArchivistaBox has **fully automated import functions**. In order to carry out a transfer, the data has to be exported from the old system. You can then create an import file and then import it to the ArchivistaBox by using an USB stick ('transfer' folder).

🔄: You can easily identify the format to be used by exporting two or three files from the ArchivistaBox or ArchivistaGPL.

33 Solutions for common problems

33.1 Scanning is slow

Scanning may be slowed down by some of the options. Functions that use considerable calculating capacity include:

- OCR recognition
- PDF generation
- Barcode recognition (optional module)
- Removal of blank pages

If several of these are turned on at the same time and the system is supposed to go through all tasks at once scanning speed may sink. If you want all of these functions, turn them off during scanning and turn them back on afterwards.

33.2 Scanner is not working

The first question you should ask yourself is: am I using the right scan definition? If your papers are sitting in the ADF but you are using a scan definition where 'ADF' as option is not chosen, then the scanner cannot respond.

The second possibility may be that the connection between ArchivistaBox and scanner is broken.

A scanner with USB interface should immediately respond when a connection with the mains on the one hand and the ArchivistaBox on the other has been established. A startup procedure where the scanner has to be turned on first and the ArchivistaBox afterwards is not required. However, sometimes this connection is interrupted

without any apparent reason. In such a case simply turn the scanner off, and after a minute turn it back on. In 9 out of 10 cases the connection will be resumed. If not, try as a last resort shutting down the ArchivistaBox and starting it up again. See 24.10 und 5.2.

33.3 Characters on keyboard do not work

If you cannot type characters that are represented on the keys of your keyboard, then you must adjust the keyboard layout setting. See above in 25.14.1.1.

33.4 OCR is not (or no longer) functioning

Should the text recognition not be working, the underlying causes may be the following:

- OCR module hangs
- OCR has not been activated for this particular database
- A 'warped' log entry prevents the OCR software from working

The measures below should put things right (order according to the causes listed above):

- Restart ArchivistaBox (gives you a fresh OCR installation)
- OCR must be turned on for the database in question; see 14.1
- Delete log table and restart OCR; see 25.14.4.3 and 25.14.4.5

33.5 ArchivistaBox is no longer working

If the box does not function, please contact the vendor from whom you bought it, or Archivista GmbH. You will get a free replacement if the ArchivistaBox hardware is damaged.

33.6 A user has forgotten their password

You can reset a user's password via WebConfig. See the information under 24.7.

33.7 'Document ... locked' message is displayed

If you do not exit the program properly (i.e. by pressing Ctrl+Alt+Del to close the application), a document that is undergoing processing may be locked. You can unlock documents that are locked in this way via WebConfig under 'Unlock documents'. See the information under 24.6.

Index

Address of Archivista GmbH, 9

API

Applications, 187

Archivista tables, 183

Login to databases, 187

Applications

own ones, 187

Archive - Introduction

folder, 23

pages, 23

Archive administration - WebAdmin, 65

Archive, digital - Introduction, 20

Archiving concept, 183

Archiving process - ArchivistaBox, 108

Archivista document, the - Introduction, 23

Archivista-BarcodePrint, 78

ArchivistaBox, 19, 25

ArchivistaBox Cloud, 19

ArchivistaBox Mini, 19

ArchivistaBox:Albis, 32

ArchivistaBox:Dolder, 30

ArchivistaBox:ISO, 20

ArchivistaCloud, 19

ArchivistaDMS, 19

ArchivistaDMS:Cloud, 19

ArchivistaMini, 19

ArchivistaVM, 19, 135

ArchivistaVM:Mini, 19

Auditing acceptability - ArchivistaBox, 192

Backup, 114

Barcode

Code 25, 78

Code 39, 78

Interleaved 2 of 5, 78

Barcode processing - RichClient, 77

Why?, 77

Barcode technology - RichClient, 77

Barcodes - WebAdmin, 77

Blanking fields - WebAdmin, 64

Client - Introduction, 21

Computer files - Publishing Edition, 165

Controlled vocabulary - WebClient, 40

Conversion, 199

Copyright, 14

Creating and scanning documents - WebClient, 44

Database - Introduction, 21

Database conversion, 199

Database creation - WebAdmin, 100

Digital archive - Introduction, 20
DMS - Introduction, 22
Documents - Introduction, 22
Dolder, 30
Download - WebClient, 49

E-mails, archiving of - ArchivistaBox, 121
Edit mode - WebClient, 40
Editing and deleting pages - WebClient, 46
Electronic documents, archiving of - ArchivistaBox, 117
Electronic files - Publishing Edition, 165
Encryption - ArchivistaBox, 116

FAQ - ArchivistaBox, 191
Field definition - WebAdmin, 60
Field names, 188
Field types - WebClient, 40
Fields - Introduction, 21
File upload, 120
File upload - WebClient, 44
Filing with simple categories - Introduction, 22
Folders, 186
Folders - Introduction, 21
FTP-Server - ArchivistaBox, 119
Full text retrieval archive - Introduction, 22
full text search, 9
Function keys - ArchivistaBox, 132

Help, 8

Icons - WebClient, 36
Installation - Introduction, 16
Introduction, 8
ISO file, 20

KVM, 134

Login - WebClient, 35
Long-term considerations on documents - Publishing Edition, 160

Mail archiving - WebAdmin, 97
Mail server - ArchivistaBox, 121
Main view and page view - WebClient, 35
Manual - Introduction, 8
Manual - WebConfig, 102
Manuals - ArchivistaBox, 107
Mask definition - WebAdmin, 61
Master-Slave configuration - ArchivistaBox, 121
Modules - ArchivistaBox, 111

Notes, 46

OCR definitions - WebAdmin, 86
Office files, 120
online help, 8
Online-Update, 131
Operating system, 20

Pages - Introduction, 22
Passwords - ArchivistaBox, 130
PDF file, creation of - WebClient, 44
PDF files - Publishing Edition, 165
Preliminary note - Publishing Edition, 162
Printing pages - WebClient, 49
Publishing - WebClient, 42
PublishingEdition, 160

RAM mode, 20
Record - Introduction, 21

Remote access - ArchivistaBox, 131
Restore-on-the-fly, 116

Saving your data - ArchivistaBox, 192
Scan definitions - WebAdmin, 72
Scanning with network scanner - ArchivistaBox, 119
Searching - WebClient, 38
Self-made applications, 187
Server - Introduction, 21
Settings - ArchivistaBox, 107
Shut down - ArchivistaBox, 132
Source formats - Publishing Edition, 161
SQL definitions - WebAdmin, 88
Startup - ArchivistaBox, 25
Support, 14
System - ArchivistaBox, 126

Tables - Introduction, 21
Template, copying of - WebClient, 44
Transferring a database to the ArchivistaBox, 199
Tutorial, 160

Update, 131
Update ArchivistaBox, 28
User administration - WebAdmin, 55
User manual - WebClient, 35

View - WebClient, 38
Virtualisation, 134
Virtualisierung, 20
Virtualisierung:Mini, 19

Web pages - Publishing Edition, 166
WebAdmin - ArchivistaBox, 107
WebClient - ArchivistaBox, 107
WinUpload, 178
Workflow module, 112
Working method - Introduction, 21