# Getting Started with AutoPortfolio<sup>™</sup>

# Plug-in for Adobe<sup>®</sup> Acrobat<sup>®</sup>

Starter Guide to Converting and Processing PDF Portfolios

# **Contents**

What is AutoPortfolio?	2
Converting a PDF portfolio into a single, "flat" PDF document	2
Extracting all files from a PDF portfolio	2
Extracting PDF Portfolio metadata into Excel spreadsheet	2
Converting a PDF Portfolio into TIFF image and text format	2
De-duplicating PDF and load files	2
Converting a PDF Portfolio into a single PDF document	3
Extracting All Files from a PDF Portfolio	. 12
Extracting PDF Portfolio metadata into Microsoft Excel Spreadsheet	. 18
Exporting PDF Portfolio into TIFF and Text format	. 21

# What is AutoPortfolio?

The AutoPortfolio is a plug-in (add-on) for Adobe<sup>®</sup> Acrobat<sup>®</sup> software. It is designed to perform the following operations:

# Converting a PDF portfolio into a single, "flat" PDF document

- All portfolio items/documents are placed into a single PDF document.
- File attachments can be optionally converted to PDF and placed right after parent documents.
- MSG and ZIP files are automatically extracted and their contents processed.
- The ability to process entries only for a selected date or a person (as in the case of email).
- The output document is hierarchically bookmarked and is ready to be Bates stamped.
- Portfolio metadata is exported into a number of spreadsheet ready formats.

# **Extracting all files from PDF portfolio**

- All items/documents are extracted as separated files into a single output folder.
- File attachments can be optionally converted to PDF and appended to parent documents.
- The ability to process entries only for a selected date or a person (as in the case of email).
- The contents of MSG and ZIP files are automatically processed.
- Portfolio metadata is exported into a number of spreadsheet ready formats.

# **Extracting PDF Portfolio metadata into an Excel spreadsheet**

- All metadata information is extracted into an Excel spreadsheet (XML or text CSV formats). If
  processing email portfolio this includes all email fields such as "From", "To", "Subject" and etc.
- The ability to extract information only for a selected date or a person (as in the case of email).
- MD5 checksums are automatically computed for all files.

# **Converting a PDF Portfolio into TIFF image and text format**

- Output is suitable for importing into litigation support systems such as Concordance and Summation.
- The ability to process entries only for a selected date or a person (as in the case of email).

# **De-duplicating PDF and load files**

- Finding duplicate and near duplicate PDF documents.
- Specifically designed for comparing and de-duplicating emails.
- Significantly reduces the amount of documents that need to printed or processed.

All operations are available from the main Acrobat menu via the "Plug-ins > AutoPortfolio" menu.

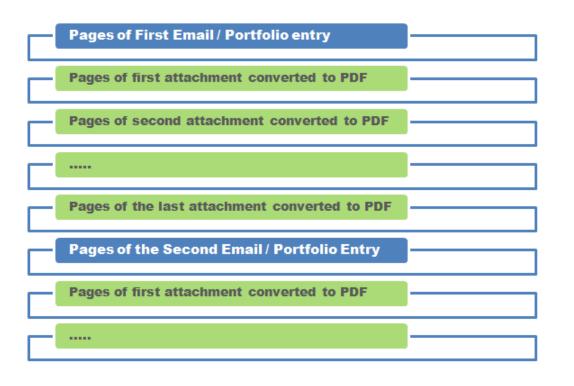
Plug-Ins Window Help	
AutoPortfolio Plug-in 🔹 🕨	Documentation and Getting Started
	Convert Portfolios Into Regular PDF
	Convert PDF Files For Concordance and Summation (TIFF and Text)
	Extract Files From Portfolio(s)
	Extract Portfolio Metadata
	Deduplicate Load Files

Do not open PDF Portfolio directly in Adobe Acrobat – use the plug-in menu to start AutoPortfolio. If a PDF Portfolio is open in Adobe Acrobat, then the "Plug-ins" menu is automatically disabled. Every operation provides a way to select one or more input PDF portfolios.

# **Converting a PDF Portfolio into a single PDF document**

This operation takes one or more PDF Portfolios and converts them into a regular "flat" PDF document. All portfolio attachments can be optionally converted to PDF format and appended right after their parent document. Every document and attachment is hierarchically bookmarked in the output PDF document. Metadata and page index file is exported into a number of spreadsheet-ready formats. If an attachment cannot be converted into PDF format, a stub page is generated and inserted into output.

Here is a logical structure of the output PDF document for a sample PDF Portfolio that contains emails:



# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

Select "Plug-ins > Convert Portfolios into Regular PDF" from the main Adobe Acrobat menu to open a conversion dialog (do not open a portfolio directly in Acrobat):

Convert PDF Portfol	ios Into Regular PDF	×
	his dialog to convert embedded files from one or mor hments) into a single regular PDF file.	e PDF Portfolios (including
Select Input Portf	olios and Regular PDF Files	
Process eac	h portfolio into a separate output (create output sub	p-folders)
Select Output Opt	ions	Add Files       Remove       Remove All       Sort Order
Output folder:	C:\Data\portfolio\	Browse
Convert non	process attachments -PDF attachments into PDF format (if possible) n output documents	Conversion Options File Naming Options
Help	www.evermap.com	OK Cancel

Press the "Add Files..." button to select the input PDF portfolio for processing; the "Specify Sort Order" dialog will appear on screen. This dialog allows for the selection of only a part of the PDF Portfolio for processing either by checking a box in front of every record or by performing a text search. If you want to select only a specific subset of the portfolio entries for processing, then use the "Select Records" menu to manipulate the current selection set.

Specify Sorting Order - 4 record(s) Select Records Click on column headers to sort records into desired processing order. Use "Select Records" menu for more options. Description Filename  $\checkmark$ THIS IS A PORTFOLIO TEST.pdf "EverMap LLC" 9/12/2011 THIS IS A PORTFOL ☑ Test for JPG attachments.pdf "EverMap LLC" 9/13/2011 Test for JPG attachr  $\checkmark$ Test for MS Word attachment.pdf "EverMap LLC" 9/13/2011 Test for MS Word a  $\checkmark$ Text File Attachment Example.pdf "EverMap LLC" 9/13/2011 Text File Attachmer

If you want to process the entire portfolio, then simply press the "OK" button located in the lower-right corner of the screen. Select an output folder by pressing the "Browse..." button:

Convert PDF Portfol	Convert PDF Portfolios Into Regular PDF							
Information Use ti attact	his dialog to convert embedded files from one or mo hments) into a single regular PDF file.	re PDF Portfolios (including						
	olios and Regular PDF Files h portfolio into a separate output (create output su	b-folders)						
C: EverMap PI	DF\Test Documents\AutoPortfolio\Emails With Differ	Add Files Remove Remove All Sort Order						
Convert non	ions C:\Data\portfolio\ process attachments -PDF attachments into PDF format (if possible) n output documents	Browse Conversion Options File Naming Options						
Help	www.evermap.com	OK Cancel						

It is recommended that you select a short output path (like c:\Data\) located on the local hard drive for better performance.

Check the "Extract and process attachments" and "Convert non-PDF attachments into PDF format" options if you want to process file attachments for each portfolio entry. For example, if a PDF Portfolio was created by exporting emails from Microsoft Outlook, then these options will convert email attachments into PDF format and append them right after the parent email. If a certain attachment cannot be converted into PDF format (for example, a file is password-protected or is not in a file format that can be converted into PDF), then a stub-page will be inserted into the output file stating this fact. The stub page will show the attachment name and file size. Every attachment is hierarchically bookmarked for easy navigation and preservation of the original document structure.

Check the "Delete links in output documents" option if you want to completely remove the file links to file attachments in their native file formats as they are originally stored within the PDF Portfolio. For example, if an email contains a Microsoft Word attachment, then in the output file, the link to the MS

## [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

Word document will be removed. The attachment will only be available in its PDF representation (if the "Extract and process attachments" option is selected). The output file size may be significantly decreased if this option is used, since all attachments are not going to be stored twice (in the native and PDF formats).

Click the "Conversion Options..." button to open the "File Conversion Options" dialog.

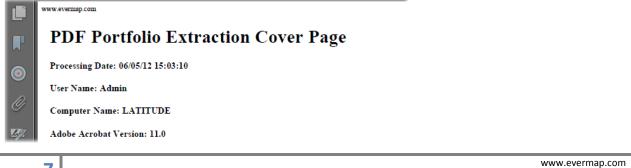
File extensions (type one per line):	Convert unknown file types as plain text Extract content of ZIP and MSG files
~	<ul> <li>Insert a cover page in front of each document</li> <li>Cover page style: Name of the File or Attachment</li> <li>Copy files that did not convert to a folder: Select Folder</li> </ul>
TIP: type one file extension per line without a period symbol. For example: "WMV"	Output folder: C:\Data\Files That Did Not Convert

This dialog provides a way to select the options that affect how certain non-PDF files are converted into PDF format. Click the "Select Folder..." button and select a folder on your computer where you want to place copies of the file attachments that did not convert into PDF format or PDF files that cannot be inserted into output due to security restrictions. This option is useful for detailed inspection of the conversion results and provides a way to preserve documents that cannot be converted into PDF format. Press" OK" to close the "File Conversion Options" dialog.

Now, everything is ready to actually run a conversion from a PDF Portfolio into a single, regular PDF document. Press "OK" to start processing:

Convert PDF Portfoli	ios Into Regular PDF	×
	his dialog to convert embedded files from one or mo hments) into a single regular PDF file.	re PDF Portfolios (including
Select Input Portfo	olios and Regular PDF Files	
Process ead	h portfolio into a separate output (create output su	b-folders)
	DF\Test Documents\AutoPortfolio\Emails With Differ	Add Files Remove Remove All
Select Output Opt Output folder:	ions C:\Data\portfolio\	Sort Order
Convert non	process attachments -PDF attachments into PDF format (if possible) n output documents	Conversion Options File Naming Options
Help	www.evermap.com	OK Cancel

Please note that the conversion process may take a considerable amount of time depending on the size of the input portfolio. It is generally a good idea to process smaller portfolios. The "PDF Portfolio Extraction Cover Page" document is automatically created for each job and displayed on the screen. The standard Acrobat progress dialog shows the progress at the bottom-right corner of the screen.



# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

Once the processing is completed, a report message is going to appear on the screen asking if you want to display a processing report. Click "OK" to display a detailed processing report in your default web browser (this report is in HTML format):

Adobe A	crobat X
1	8 file(s) have been extracted from 1 PDF portfolio(s). Do you want to display processing report?
	OK Cancel

Here is a sample processing report that lists processed files (there is also spreadsheet ready version of the report in CSV file format). The report lists the file name, description (from a metadata field), creation and modification dates, file size in bytes, number of attachments, and MD5 hash value:

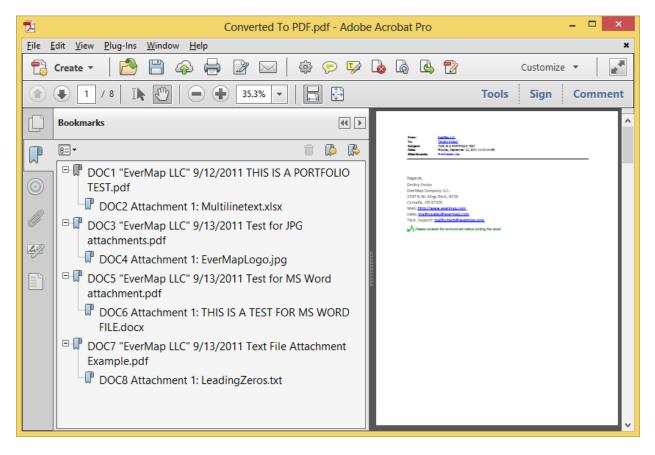
🕨 Ϲ 👚 🕓 file	e:///C:/Data/portfo	lio/PDF%20Portfo	lio%20Ex	traction%	20Repo	ort.htm	🖾 🖧 🐻
Number of PDI ut folder: C:\Da adsheet-ready ( Input PDF Package	F Portfolios proc ta\portfolio\ CSV text file) ve / File: <u>C:\EverMap P</u> racted files (first level	rsion of this rep	ort: <u>C:\D</u>	ata\port		DF Portfolio Extraction Report.	<u>csv</u>
Number of extracted attachments: 4					Num Attachment		
THIS IS A PORTFOLIO TEST.pdf	THIS IS A PORTFOLIO TEST.pdf	"EverMap LLC" 9/12/2011 THIS IS A PORTFOLIO TEST.pdf	09/13/11 08:29:43		59038	edca43ddeb35b3809a5de54790dfe22b	1
Attachments:							
Attachments:			00/05/40	06/05/12		b0b1a51e23bb714ac84967e7363c7cc9	0
Attachments: Multilinetext.xlsx	Multilinetext.xlsx		06/05/12 15:03:32		8334		
	Multilinetext.xlsx Multilinetext.xlsx			15:03:32 06/05/12	8334 26351	e76d30360bf6ccf2f99bcb745df6bdb5	0

It is a good idea to inspect the report and see if there are any file attachments that were not converted into a PDF file format. You would see a red line in the table for every file attachment that failed to convert. Scroll down to the end of the "AutoPortfolio Process Report" to see the total count of non-PDF file attachments that were converted and not-converted into a PDF format.

# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

Example.pdf	Example.pdf	Attachment Example.pdf	08:29:43	08:29:43	54217	800100200 100060000111 081 5000505 1	<b>'</b>
Attachments:							
LeadingZeros.txt	LeadingZeros.txt		06/05/12 15:04:35			a891b28275485120d2afe0f041f2acf3	0
LeadingZeros.pdf	LeadingZeros.txt		06/05/12 15:04:35		4214	5f7084a56120c17417376f9dd318c67c	0

Now, let's look at the output PDF document that is going to be automatically open in Adobe Acrobat once processing is completed:



All individual PDF portfolio entries (PDF files) are now stored in a single PDF document that can be processed using standard Adobe Acrobat tools. Every file (email in the case of email portfolios) is bookmarked with child bookmarks pointing to the corresponding file attachments. The "Description" metadata field and unique document ID is used for the bookmark title. Click on the bookmark to automatically display a corresponding document or attachment. The output document is ready now to

be Bates-stamped and/or printed. The output document is automatically named "Converted To PDF.pdf". Use the "File > Save As..." menu to save this document under a different filename or location.

Now, let's look at the output folder and list all the files that are created after the processing. Here is an output from a sample PDF portfolio named *Emails With Different Attachments.pdf*:

🔁 Converted To PDF.pdf

- Converted To PDF.pdf.DocPageIndex.txt
- 😻 CoverPage.htm
- 🔁 CoverPage.pdf
- 🔨 Emails With Different Attachments.pdf.report.csv
- 🐏 Emails With Different Attachments.pdf.report.xml
- BDF Portfolio Extraction Report.csv
- PDF Portfolio Extraction Report.htm

The output folder contains the following output files:

- Converted To PDF.pdf output PDF file that contains all entries from the input portfolio as a single, "flat" PDF document.
- *Converted To PDF.pdf.DocPageIndex.txt* a tab-delimited plain text file that maps document ID to a page number in the output document.
- *CoverPage.pdf* cover page document that contains information about who and when ran the processing.
- CoverPage.htm same as the above, but in HTML format
- NamedOfTheInputPortfolio.pdf.report.csv Plain text CSV spreadsheet (can be opened directly in Microsoft Excel) that contains all metadata for all PDF Portfolio entries. In the case of the email portfolio, this file will contain "To", "From", "Subject" and other similar fields. This file also lists document IDs that used for bookmarking and in the page index file.
- PDF Portfolio Extraction Report.csv plain text CSV spreadsheet (can be opened directly in Microsoft Excel) that contains a processing report listing all files that have been processed and converted.
- *PDF Portfolio Extraction Report.htm* HTML version of the processing report.

Here is an example of Converted To PDF.pdf.DocPageIndex.txt page index file:

```
DOC1
        "EverMap LLC" 9/12/2011 THIS IS A PORTFOLIO TEST.pdf
                                                                1
DOC2
       Attachment 1: Multilinetext.xlsx
                                                2
DOC3
        "EverMap LLC" 9/13/2011 Test for JPG attachments.pdf
                                                                3
DOC4
        Attachment 1: EverMapLogo.jpg
                                        4
DOC5
        "EverMap LLC" 9/13/2011 Test for MS Word attachment.pdf 5
D0C6
        Attachment 1: THIS IS A TEST FOR MS WORD FILE.docx
DOC7
        "EverMap LLC" 9/13/2011 Text File Attachment Example.pdf
                                                                         7
D0C8
       Attachment 1: LeadingZeros.txt 8
```

# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]



IMPORTANT: Please note that all file format conversions are done by Adobe Acrobat itself using currently selected preferences. Select "File > Preferences..." from the menu to edit the preferences. Select the "Convert to PDF" category, then select a file format to edit and click

the "Edit Settings..." button. You may want to make changes to the default settings for Microsoft Excel conversion to allow processing of all worksheets in the file. By default, Adobe Acrobat will only convert the first worksheet in the spreadsheet.



There are certain kinds of PDF and Microsoft Office documents that cannot be merged by Adobe Acrobat into a single PDF document:

- - Password-protected PDF documents (you need to remove passwords before merging)
  - Password-protected or encrypted MS Office files
  - PDF documents with security restrictions that prevent document merging
  - PDF documents with security restrictions that do not allow printng
  - MS Office files with access/editing restrictions
  - Certain PDF forms created by Adobe LiveCycle Designer form editor •

# **Extract All Files from PDF Portfolio**

Select the "Plug-ins > AutoPortfolio Plug-in > Extract Files from Portfolio" menu to extract all files (including attachments) from one or more PDF Portfolios. Do not open a PDF Portfolio directly in Adobe Acrobat; otherwise Acrobat will automatically disable most menus including the "Plug-ins".

Extract Files From Pl	DF Portfolio	×
Information Use the Portfo	his dialog to extract all embedded files and their a blios.	attachments from one or more PDF
	olios and Regular PDF Files h portfolio into a separate output (create output	sub falders)
		Add Files Remove Remove All Sort Order
Convert non	ions C:\Data\portfolio\ process attachments -PDF attachments into PDF format (if possible) n output documents chments to parent document	Browse Conversion Options File Naming Options
Help	www.evermap.com	OK Cancel

Click the "Add Files..." button to select the input PDF portfolio for processing. The "Specify Sorting Order" dialog will appear on screen once the input file is selected:

1	Specify Sorting Order - 4 record(s)					
<	elect	Records				
	Click on column headers to sort records into desired processing order. Use "Select Records" menu for more options.					
		Filename	Description			

	$\wedge$	Filename	Description
		THIS IS A PORTFOLIO TEST.pdf	"EverMap LLC" 9/12/2011 THIS IS A PORTFOL
	☑	Test for JPG attachments.pdf	"EverMap LLC" 9/13/2011 Test for JPG attach
	☑	Test for MS Word attachment.pdf	"EverMap LLC" 9/13/2011 Test for MS Word a
		Text File Attachment Example.pdf	"EverMap LLC" 9/13/2011 Text File Attachmer
- 1			

www.evermap.com

This dialog allows selecting only a part of the PDF Portfolio for processing either by checking a box in front of every record or by performing a text search. If you want to process the entire portfolio, then simply press the "OK" button located in the bottom-right corner of the screen.

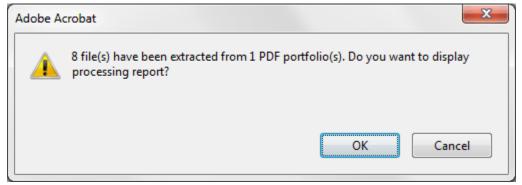
Select an output folder by pressing the "Browse…" button. Check "Extract and process attachments" if you want to extract all attachment files as well. Check the "Convert non-PDF attachments into PDF format" option if you want to convert all supported non-PDF files into PDF format. Check "Append attachments to parent documents" if you want to append all file attachments at the end of their parent PDF documents. This way each portfolio entry ("email" in case of email-based portfolios) contains all its attachments in a single PDF file.

tract Files From P	DF Portfolio			×
Information	his dialog to extract all embedde	d files and their att	achmanta from one	or more PDE
	olios.	u nies and their att	actiments from one	
Select Input Port	olios and Regular PDF Files			
Process ead	h portfolio into a separate outpu	t (create output su	ıb-folders)	
C:\EverMap P	DF\Test Documents\AutoPortfolio	o\Emails With Differ	ent Attachm	Add Files Remove Remove All
- Select Output Op	tions			Sort Order
Output folder:	C: \Data \portfolio \			Browse
Convert nor	process attachments n-PDF attachments into PDF form in output documents achments to parent document	at (if possible)	Conversion File Naming	Options
Help	www.eve	ermap.com	ОК	Cancel

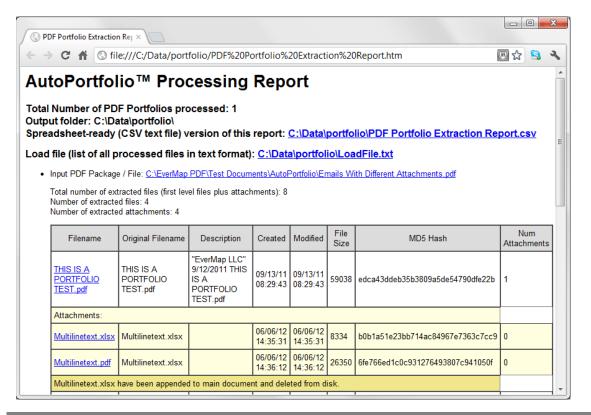
Press "OK" to start extracting files from the input PDF portfolio. The "PDF Portfolio Extraction Cover Page" document is automatically created for each job and displayed on the screen. The standard Acrobat progress dialog shows the progress at the bottom-right corner of the screen.

# www.evermap.com PDF Portfolio Extraction Cover Page Processing Date: 06/05/12 15:03:10 User Name: Admin Computer Name: LATITUDE Adobe Acrobat Version: 11.0

Once processing is completed, the report message will appear on the screen asking if you want to display a processing report. Click "OK" to display a detailed processing report in your default web browser (this report is in HTML format):



Here is a sample processing report that lists processed files (there is also a spreadsheet ready version of the report in CSV file format). The report lists file name, description (from a metadata field), creation and modification dates, file size in bytes, number of attachments, and MD5 hash value.



# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

The processing report lists every file and attachment that was processed and contains separate records for the attachments that were converted into PDF format and optionally appended to the parent document. It is a good idea to inspect the report and see if there are any file attachments that were not converted into the PDF file format. You would see a red line in the table for every file attachment that failed to convert. Scroll down to the end of the "AutoPortfolio Process Report" to see the total count of non-PDF file attachments that were converted and not-converted into a PDF format.

		e:///C:/Data/portfolio/P				DReport.htm 🛛 🕅 🏠 🦌
Leading	Zeros.txt	LeadingZeros.txt		06/06/12 14:36:29	30	a891b28275485120d2afe0f041f2ac
Leading	<u>Zeros.pdf</u>	LeadingZeros.txt		06/06/12 14:36:29	4214	7b3a4850061b3a0affed1a9c74792b
Leading	Zeros.txt I	nave been appended to mair	document and dele	ted from di	isk.	

Now, let's take a look at the content of the output folder:

😋 🔵 🗢 🚺 🕨 Comp	puter	► OS (C:) ► Data ► portfolio	Search portfolio	م
<u>F</u> ile <u>E</u> dit <u>V</u> iew <u>T</u> o	ools	<u>H</u> elp		
Organize 👻 Includ	le in li	ibrary  Share with  New folder		:=
🔆 Favorites	<b>^</b>	Name	Date modified	Туре
属 Downloads		CoverPage.htm	6/6/2012 2:35 PM	Chrome HTML Do
🖳 Recent Places	=	🔁 CoverPage.pdf	6/6/2012 2:35 PM	Adobe Acrobat D
🧾 Desktop	-	📳 Emails With Different Attachments.pdf.report.csv	6/6/2012 2:36 PM	Microsoft Office E
		LoadFile.txt	6/6/2012 2:36 PM	Text Document
🧾 Desktop		🖼 PDF Portfolio Extraction Report.csv	6/6/2012 2:36 PM	Microsoft Office E
🤭 Libraries		PDF Portfolio Extraction Report.htm	6/6/2012 2:36 PM	Chrome HTML Do
Documents		🔁 Test for JPG attachments.pdf	6/6/2012 2:36 PM	Adobe Acrobat D
🌙 Music		🔁 Test for MS Word attachment.pdf	6/6/2012 2:36 PM	Adobe Acrobat D
📔 Pictures		🔁 Text File Attachment Example.pdf	6/6/2012 2:36 PM	Adobe Acrobat D
📑 Videos		🔁 THIS IS A PORTFOLIO TEST.pdf	6/6/2012 2:36 PM	Adobe Acrobat D
輚 Homegroup	+ -	( III		•
10 items				

The above snapshot of the output folder is the result of processing when the following processing options were selected:

Select Output Opti	ons	
Output folder:	C:\Data\portfolio\	Browse
Convert non-	rocess attachments PDF attachments into PDF format (if possible) output documents hments to parent document	Conversion Options File Naming Options

The output folder does not contain attachments in their native file formats since the option "Convert non-PDF attachments into PDF format" was used. The below is another snapshot of the output folder when this option was turned off (as well as the "Append attachments to parent documents"):

File Edit View Tools	Help			ACREASE AND
Organize <b>v</b> Include in			III •	
	Name	Date modified		Size
쑦 Favorites 👘	Name	Date modified	Туре	SIZE
🐌 Downloads	CoverPage.htm	6/6/2012 2:58 PM	Chrome HTML Do	11
🖳 Recent Places 😑	🔁 CoverPage.pdf	6/6/2012 2:58 PM	Adobe Acrobat D	17 8
🧾 Desktop	鸖 Emails With Different Attachments.pdf.report.csv	6/6/2012 2:58 PM	Microsoft Office E	3 k
	🔤 EverMapLogo.jpg	6/6/2012 2:58 PM	JPG File	31
🗾 Desktop	LeadingZeros.txt	6/6/2012 2:58 PM	Text Document	11
🥞 Libraries	LoadFile.txt	6/6/2012 2:58 PM	Text Document	11
Documents	🗐 Multilinetext.xlsx	6/6/2012 2:58 PM	Microsoft Office E	91
J Music	🚳 PDF Portfolio Extraction Report.csv	6/6/2012 2:58 PM	Microsoft Office E	3 k
Pictures	PDF Portfolio Extraction Report.htm	6/6/2012 2:58 PM	Chrome HTML Do	51
📑 Videos	🔁 Test for JPG attachments.pdf	6/6/2012 2:58 PM	Adobe Acrobat D	55 H
🔞 Homegroup	🔁 Test for MS Word attachment.pdf	6/6/2012 2:58 PM	Adobe Acrobat D	60 k
🥦 Admin	🔁 Text File Attachment Example.pdf	6/6/2012 2:58 PM	Adobe Acrobat D	53 k
💻 Computer	THIS IS A PORTFOLIO TEST.pdf	6/6/2012 2:58 PM	Adobe Acrobat D	58 k
🏭 OS (C:)	THIS IS A TEST FOR MS WORD FILE.docx	6/6/2012 2:58 PM	Microsoft Office	10 1
🍶 Art 🖕				
14 items	< III			

The output folder now contains file attachments as well as PDF portfolio items in their native formats. All main level PDF Portfolio items are exported as PDF files (they are already stored in the portfolio in PDF format). The output folder also contains a number of various auxiliary files such as extraction reports (both in CSV and HTML formats), processing cover page, and \*.pdf.report.csv file (can be opened directly by Microsoft Excel) that contains metadata for all extracted files. There is also a file called

LoadFile.txt (in CaseMap format) that lists all the extracted top-level files (in the case of email-based portfolios only top level emails):

DeadFile.txt - Notepad	X
<u>File Edit Format View H</u> elp	
Linked File, Bates - Begin C:\Data\portfolio\THIS IS A PORTFOLIO TEST.pdf, THIS IS A PORTFOLIO TE C:\Data\portfolio\Test for JPG attachments.pdf, Test for JPG attachments C:\Data\portfolio\Test for MS Word attachment.pdf, Test for MS Word attachment C:\Data\portfolio\Test File Attachment Example.pdf, Text File Attachment Example	t

If the input PDF Portfolio contains an MSG file (Microsoft Outlook message format) or ZIP archive, then a folder is created for each file that contains files from the corresponding MSG and ZIP archives. If there are nested MSG or ZIP files, the second-level subfolders are automatically created and the files extracted.

# **Extracting PDF Portfolio metadata into Microsoft Excel Spreadsheet**

The PDF Portfolio is a single storage file that contains a number of files along with the corresponding metadata associated with each file. The actual metadata fields included totally depend on the application that created the particular portfolio. The most common use of PDF Portfolio is to store emails exported from email applications such as Microsoft Outlook or Lotus Notes. In this case, most metadata fields come from the corresponding email's metadata and typically include "To", "From", "Subject" and etc. The AutoPortfolio plug-in can export all this information into an Excel spreadsheet file. The plug-in exports <u>all fields</u> that are actually present in the PDF Portfolio along with some additional metadata fields that are computed by the plug-in itself (MD5 hash).

Information		
Use	this dialog to extract metadata associated with fil folios. Metadata is exported in either Microsoft Ex /).	
	folios and Regular PDF Files	1
Process ea	ch portfolio into a separate output (create output	t sub-folders)
		Remove Remove All
Select Output Op	otions	
Select Output Op Output folder:		Browse
		Browse Conversion Options File Naming Options

Click the "Add Files..." button to select the input PDF Portfolio and "OK" to start processing.

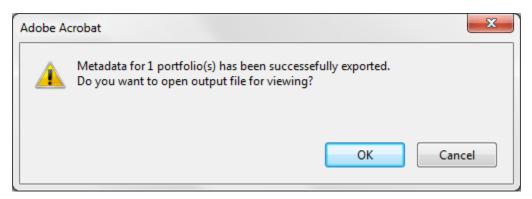
#### [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

🔁 Save As				X
Save <u>i</u> n:	鷆 portfolio	•	🎯 🤌 📂 🛄 <del>-</del>	
Ca	Name	*	Date modified	Туре
Recent Places		No items match your s	earch.	
Desktop				
Libraries				
Computer				
() Network	•	m		۶.
	File <u>n</u> ame:	Emails With Different Attachments x	ml 🔻	<u>S</u> ave
(	Save as type:	Microsoft Excel XML File (* xml) Microsoft Excel XML File (* xml)		Cancel

Use the "Select Type" option menu in the "Save As" dialog to choose a desired output file format:

The plug-in supports two output formats: Microsoft Excel XML format and plain text ASCII \*.CSV file format. Both output formats contain the same data.

Press "OK" on the final confirmation dialog to open the output file in Microsoft Excel:



A typical email-based portfolio exported from Microsoft Outlook has the following metadata fields for each email entry:

- "Portfolio" the full path to the portfolio .
- "Filename" original filename associated with the email entry.
- "Description" a composite field that combines "From" field with data and email subject.

# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

- "Created/Modified" the date this entry was created and modified (this is not the "Sent" date).
- "Size" the actual size in bytes of the email entry (typically the size of a corresponding PDF file).
- "MD5 Hash" MD5 hash value for the corresponding file. This number is unique and can be used to compare files.
- "Folder location" the name of the email folder where this email came from (Personal Folders/Inbox for example).
- "From" the "From" field of the corresponding email message.
- "To" the "To" field of the corresponding email message.
- "GUID" globally unique identifier for the corresponding email message.
- "Date" the actual date the email was received.
- "Attachments" the number of attachments in the corresponding email message
- "Subject" subject field of the corresponding email message.

The above fields are listed as an example only. The actual output depends on the actual metadata fields stored in the specific PDF Portfolio. The plug-in does not look for any pre-defined fields; it exports all fields that are present in the actual file.

6	0		(~ - ) ∓									Emails V	Vith Different
	0	Home	Insert	Page Layout	Formulas	Data	Review	View	Acrobat	Team			
	Pa	Cut □ Copy ste ↓ Format Clipboard	Painter	Calibri B Z U ·	• 11 • A /				Wrap Text	iter -	General	▼ .00, 00, .00 ⇒.0	Conditional Formatting
F		A1		() fx	Portfolio			igninen			Number		
F		A			В			С			D		E
	1	Portfolio		Filename			Description			Cre	ated	Modifie	ed .
	2	C:\EverMap F	PDF\Test	THIS IS A PORT	FOLIO TEST.pd	lf	"EverMap Ll	.C" 9/1	2/2011 THIS	S A 09/	13/11 08:29:43	09/13/1	1 08:29:43
	3	C:\EverMap F	PDF\Test	Test for JPG at	tachments.pdf	f	"EverMap Ll	.C" 9/1	3/2011 Test f	or 09/	13/11 08:29:43	09/13/1	1 08:29:43
	4	C:\EverMap F	PDF\Test	Test for MS W	ord attachmen	t.pdf	"EverMap Ll	.C" 9/1	3/2011 Test f	or 09/	13/11 08:29:42	09/13/1	1 08:29:42
	5	C:\EverMap F	PDF\Test	Text File Attac	hment Examp	le.pdf	"EverMap Ll	.C" 9/1	3/2011 Text	File 09/	13/11 08:29:43	09/13/1	1 08:29:43

# **Exporting PDF Portfolio into TIFF and Text format**

The plug-in provides the ability to export a PDF Portfolio into a TIFF and Text format that is suitable for importing data into several litigation support systems such as Concordance and Summation. The output is a collection of files – one TIFF image file and one text file for each page of each document in the portfolio.

Select "Plug-ins > AutoPortfolio Plug-in > Convert PDF Files for Concordance and Summation (TIFF and Text" from the menu. Press the "Add Files..." button to select an input PDF Portfolio.

Convert Files For Co	ncordance and Summation	×			
	is dialog to convert PDF Portfolios and files into set of TIFF a ige) for loading into Concordance and Summation application				
_	lios and Regular PDF Files n portfolio into a separate output (create output sub-folders)	, 1			
		Add Files Remove Remove All Sort Order			
Select Output Opti	ons	2			
Output folder:	C:\Data\portfolio\	Browse			
Convert non	<ul> <li>Extract and process attachments</li> <li>Convert non-PDF attachments into PDF format (if possible)</li> <li>Append attachments to parent document</li> <li>Delete links in output documents</li> <li>Tile Naming Options</li> <li>File Numbering Options</li> </ul>				
Help	www.evermap.com	OK Cancel			

The "Specify Sorting Order" dialog will appear on the screen once the input file is selected. This dialog allows for the selection of only a part of the PDF Portfolio for processing either by checking a box in front of every record or by performing a text search. If you want to select only a specific subset of the portfolio entries for the processing, then use the "Select Records" menu to manipulate the current selection set

#### Specify Sorting Order - 4 record(s)

#### Select Records

Click on column headers to sort records into desired processing order. Use "Select Records" menu for more options.

	Filename	Description
	THIS IS A PORTFOLIO TEST.pdf	"EverMap LLC" 9/12/2011 THIS IS A PORTFOL
☑	Test for JPG attachments.pdf	"EverMap LLC" 9/13/2011 Test for JPG attach
$\square$	Test for MS Word attachment.pdf	"EverMap LLC" 9/13/2011 Test for MS Word a
	Text File Attachment Example.pdf	"EverMap LLC" 9/13/2011 Text File Attachmer

If you want to process the entire portfolio, then simply press "OK" located in the lower-right corner of the screen. Select an output folder by pressing the "Browse..." button. Next, press the "File Numbering Options..." button to choose the desired file numbering settings.

File Numbering Options			
File numbering options are used for labeling files when generating a Summation load file (*.dii).			
Typically, a Bates number is used to label each document entry in DII load file. Use controls below to create a file numbering scheme that matches your database.			
Options			
Number of digits:	6		
Starting number:	1		
Prefix:	A		
Suffix:			
Help	OK Cancel		

This dialog allows you to configure the plug-in to number output files to match file naming in the Summation database. Please refer to Summation documentation for details on file naming and DII import or press "Help..." to read more about these options.

Now, choose any other output option, then press "OK" to start processing. Please note that the conversion process may take a considerable amount of time depending on the size of the input portfolio. It is generally a good idea to process smaller portfolios. The processing consists of extracting every PDF document out of the portfolio and extracting any existing file attachments. Converting non-PDF attachments into PDF (optional) and then converting all resulting PDF files into TIFF/TEXT format. For every page in a PDF file there are two output files: one TIFF image and one plain text file. The "PDF Portfolio Extraction Cover Page" document is automatically created for each job and displayed on the screen. The standard Acrobat progress dialog shows the progress at the bottom-right corner of

#### the screen.

	www.evermap.com
F	PDF Portfolio Extraction Cover Page
	Processing Date: 06/05/12 15:03:10
	User Name: Admin
	Computer Name: LATITUDE
L.J.	Adobe Acrobat Version: 11.0

Once processing is completed, the report message will appear on the screen asking if you want to display a processing report. Click "OK" to display a detailed processing report in your default web browser (this report is in HTML format).

Adobe A	crobat 💌
4	8 file(s) have been extracted from 1 PDF portfolio(s). Do you want to display processing report?
	OK Cancel

#### Now, let's look at the contents of the output folder:

Comp	oute	r ► OS (C:) ► Data ► portfolio ► -	✓ Search portfolio	<mark>م</mark>
<u>File Edit View To</u>	ools	Help		
Organize 🔻 Includ	le in	library ▼ Share with ▼ New folder		:= - 🔟 🔞
Pictures	*	Name	Date modified	Туре
Videos		\mu PDF	6/7/2012 6:02 PM	File folder
Homegroup		IFF and TEXT	6/7/2012 6:02 PM	File folder
Admin	Ξ	CoverPage.htm	6/7/2012 6:01 PM	Chrome HTML Document
Computer		🔁 CoverPage.pdf	6/7/2012 6:01 PM	Adobe Acrobat Document
🕌 OS (C:)		🖼 Emails With Different Attachments.pdf.report.csv	6/7/2012 6:02 PM	Microsoft Office Excel Com
Art		LoadFile.txt	6/7/2012 6:02 PM	Text Document
AudioBooks		🖼 PDF Portfolio Extraction Report.csv	6/7/2012 6:02 PM	Microsoft Office Excel Com
Config.Msi		PDF Portfolio Extraction Report.htm	6/7/2012 6:02 PM	Chrome HTML Document
J Data				
Data2	÷	•		+
8 items				

There are two sub-folders in the output folder: "PDF" and "TIFF and TEXT". Here is an example of the content for the "PDF" folder:

						x		
😋 🔾 🗢 🚺 « OS (C	:) ▶	Data 🕨 portfolio 🕨 PDF 🛛 👻 🍫	Search PDF			م (	>	
<u>File E</u> dit <u>V</u> iew <u>T</u> o	ools	Help						
Organize 👻 Includ	le in l	ibrary ▼ Share with ▼ New folder		•		0		
Pictures	*	Name	Date modified	Туре			s	
Videos		🔁 File_1_THIS IS A PORTFOLIO TEST_Page	6/7/2012 6:02 PM	Adobe A	crobat l	D		
🔣 Homegroup		🔁 File_3_Multilinetext_Page_1.pdf	6/7/2012 6:02 PM	Adobe A	crobat l	D	.	
	=	🔁 File_4_Test for JPG attachments_Page_1.p	6/7/2012 6:02 PM	Adobe A	crobat l	D		
Computer		🔁 File_6_EverMapLogo_Page_1.pdf	6/7/2012 6:02 PM	Adobe A	crobat l	D		
Art		🔁 File_7_Test for MS Word attachment_Pag	6/7/2012 6:02 PM	Adobe A	crobat l	D		
AudioBooks		🔁 File_9_THIS IS A TEST FOR MS WORD FIL	6/7/2012 6:02 PM	Adobe A	crobat l	D		
Config.Msi		🔁 File_10_Text File Attachment Example_Pa	6/7/2012 6:02 PM	Adobe A	crobat l	D		
Data		🔁 File_12_LeadingZeros_Page_1.pdf	6/7/2012 6:02 PM	Adobe A	crobat l	D		
Data2		LoadFilePDF.txt	6/7/2012 6:02 PM	Text Doc	ument			
dell	Ŧ	< III					Þ	
9 items								

This folder contains a number of PDF files named File\_X\_\*.pdf. The total number of PDF files in this folder is equal to the total number of pages in all the documents of the input PDF Portfolio. Each PDF file contains only one page.

The "TIFF and TEXT" folder contains:

Number of \*.TXT files named File\_X\_\*.txt. There is exactly one text file for each page in the PDF Portfolio.

Number of image \*.TIF files named File\_X\_\*.tif.

Here is an example of the content for the "TIFF and TEXT" Folder:

# [GETTING STARTED WITH AUTOPORTFOLIO PLUG-IN]

		Search TIFF and T	LAT
ile <u>E</u> dit <u>V</u> iew <u>T</u> ools	<u>H</u> elp		
)rganize 🔻 👘 Include in	library ▼ Share with ▼ New folder		= •
📔 Pictures 🔷	Name	Date modified	Туре
📑 Videos	CasemapLoadFileTEXT.txt	6/7/2012 6:02 PM	Text Document
🝓 Homegroup	CasemapLoadFileTIFF.txt	6/7/2012 6:02 PM	Text Document
🦲 Admin	ConversionReport.txt	6/7/2012 6:02 PM	Text Document
🖳 Computer	ConversionReport.txt File_1_THIS IS A PORTFOLIO TEST_Page		TIF File
繿 OS (C:)	File_1_THIS IS A PORTFOLIO TEST_Page		Text Document
퉬 Art	File_3_Multilinetext_Page_1.tif	6/7/2012 6:02 PM	TIF File
🌗 AudioBooks	File_3_Multilinetext_Page_1.txt	6/7/2012 6:02 PM	Text Document
🎒 Config.Msi	File_4_Test for JPG attachments_Page_1.tif		TIF File
🌗 Data 🗏	File_4_Test for JPG attachments_Page_1.txt		Text Document
🌗 Data2	File_6_EverMapLogo_Page_1.tif	6/7/2012 6:02 PM	TIF File
🍌 dell	File_6_EverMapLogo_Page_1.txt	6/7/2012 6:02 PM	Text Document
퉬 Drivers	File_7_Test for MS Word attachment_Pag		TIF File
퉬 eBooks 💷	File_7_Test for MS Word attachment_Pag		Text Document
퉬 eBooks Otec	File_9_THIS IS A TEST FOR MS WORD FIL		TIF File
퉬 EverMap PDF	File 9 THIS IS A TEST FOR MS WORD FIL		Text Document
퉬 Intel	File_10_Text File Attachment Example_Pa		TIF File
퉬 Maps	File_10_Text File Attachment Example_Pa		Text Document
퉬 MDT	File_10_1 ext File Attachment Example_Pa File_12_LeadingZeros_Page_1.tif	6/7/2012 6:02 PM	TIF File
🚮 MSOCache	File_12_LeadingZeros_Page_1.txt	6/7/2012 6:02 PM	Text Document
퉬 PerfLogs	SummationLoadFile.dii	6/7/2012 6:02 PM	DII File
퉬 Photo 🔤		0/ 7/ 2012 0.02 FIVI	DITTIC
📜 Drogram Eiler 🍸	•		