

WP5: Tool tests for scenario 2.2

Authors:

Eva Toller (National Archives of Sweden, RA)

.....

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	P
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
0.1	20130614	Eva Toller	RA	First draft
0.2	20130618	Eva Toller	RA	Added tests results for Avis Document Converter



1 TOOL TESTING FOR SCENARIO 2.2

1.1 SCENARIO DESCRIPTION

“A university lecturer in art history wants to use a collection of digitised art images made 15 years ago. They are stored in a format he is not familiar with. Since there are about 200 images, the researcher is looking for tools which would convert them into a format he could easily use in batch mode. He is not sure how to identify a tool or a service which could do this.”

General comment: for test purposes, it may suffice to use 10 – 20 images. If this scenario is reused in Proof of Concept #2, a more realistic amount of images may be used, if possible.

Suggested test data: see document **DCH-RP_WP5_Scen-2-2_ID-66-restricted.pdf**

1.2 DISPOSITION

Chapter 2 and the following chapters are structured in the following way:

In sections X.1, a short description is given of the tool and how it works.

In sections X.2, the data set(s) that the tool will be tested on is described. If there are several data sets, they are described in sub sections: X.2.1, X.2.2, X.2.3 ... X.2.n.

In sections X.3, a summary of the results of the tests are given. If there are several data sets and the results differ significantly between them, they are described in sub-sections: X.3.1, X.3.2, X.3.3 ... X.3.n.

In sections X.4, the test results are described in more detail, and comments are given about the tool and its usability for digital cultural heritage preservation, dissemination et c.

2 AVS DOCUMENT CONVERTER 2.2

2.1 GENERAL DESCRIPTION

AVS Document Converter 2.2 converts files of **source formats**

- PDF, HTML, HTM, MHT, RTF, DOC, DOCX, ODT, PPT, PPTX, TXT, **TIFF**, TIF, EPUB, MOBI, FB2, **DjVu**, XPS

into files of **target formats**

- **PDF**, HTML, MHT, RTF, DOC, DOCX, ODT, TXT, GIF, **JPEG**, **PNG**, **TIFF**, EPUB, MOBI, FB2

It can be downloaded from <http://www.avs4you.com/AVS-Document-Converter.aspx>

Note that there is also an **AVS Image Converter** (<http://www.avs4you.com/AVS-Image-Converter.aspx>).

The reason that the document converter is tested instead is that it supports DjVu as a source format (which the Image Converter does not). However, it may later be necessary to test the AVS Image Converter too, to cover all the source and target formats (written in **bold** above).

2.2 DATA SETS

The data sets that were used to test the AVS Document Converter 2.2 are the following (for details, see: **DCH-RP_WP5_Scen-2-2_ID-66-restricted.pdf**):

- Seven TIFF images
- 106 DjVu images

The size is between 26 KB (the smallest DjVu image) and 397 MB (the largest TIFF image).

2.3 TEST RESULTS

As a whole, the behaviour of the program was not acceptable, and the required results were only partly obtained.

The results for the different types of conversions are described below. The details are described in section 2.4.

2.3.1 Loading files

It was not possible to load all seven TIFF files simultaneously, and it was altogether impossible to load the largest TIFF file. All the DjVu files could be loaded simultaneously without problems.

2.3.2 Converting TIFF files to PDF

Only three of the seven TIFF images could be successfully converted. The quality of the output was good for these files (the target files were also smaller than the source files). The largest TIFF file (that is, the largest that could be loaded) could not be converted at all. Some seemingly converted files resulted in very small target files with only a white or grey area instead of an image.

2.3.3 Converting DjVu files to PDF

Most of the files could be converted, but the larger the source file, the harder the conversion (as for the TIFF files, this resulted in very small target files with a white or grey area instead of an image). The quality of the output was good for the files that could be converted successfully. However, the target files were much larger than the source files, approximately ten times larger.

2.3.4 Converting DjVu files to JPEG

All files could be successfully converted as long as you did not attempt to convert them all at once (when this was tried, the program stopped working).

However, the quality of the output depended on the size of the source file; it was much better for the larger source files than for the smaller ones. The sizes of the target files were about two times larger than the sizes of the source files.

2.3.5 Converting DjVu files to PNG

The result was approximately the same as for the DjVu-to-JPEG conversion, except that the sizes of the target files were about four to five times larger than the source files.

2.4 USABILITY

The download and installation of the tool were quite easy and went fast (however, it was recommended that you close all other applications before installation).

A double-click on the file **AVSDocumentConverter.exe** starts the installation. The usual pop-up windows that can be expected during program installations are shown:

- A question to make certain that you trust this provider, and the provider's certificate
- Selection of a language (the choices are German, English, Spanish, French, Italian)
- Selection of the directory for installation

Each time you start the program, the following pop-up window is shown:

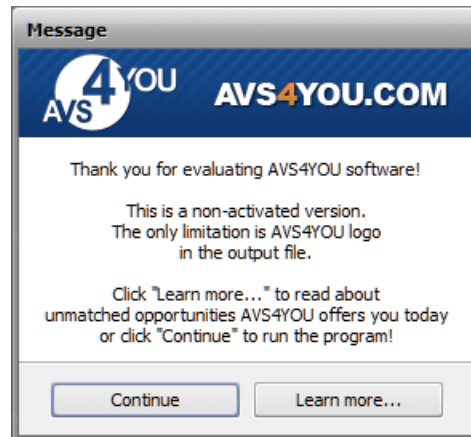


Figure 1: Pop-up window shown at program start

“Learn more” shows a web page (<http://www.avs4you.com/Register.aspx?Type=App&ProgID=74&URL=Register>) that lists the price of the commercial version (where you don't get the AVS logo, a watermark banner, in the output files). The prices are for the whole AVS toolkit (16 tools): \$59 for unlimited subscription and \$39 for one year's subscription. However, these prices are only valid until 2013-06-30.

The start page for AVS Document Converter looks like this:

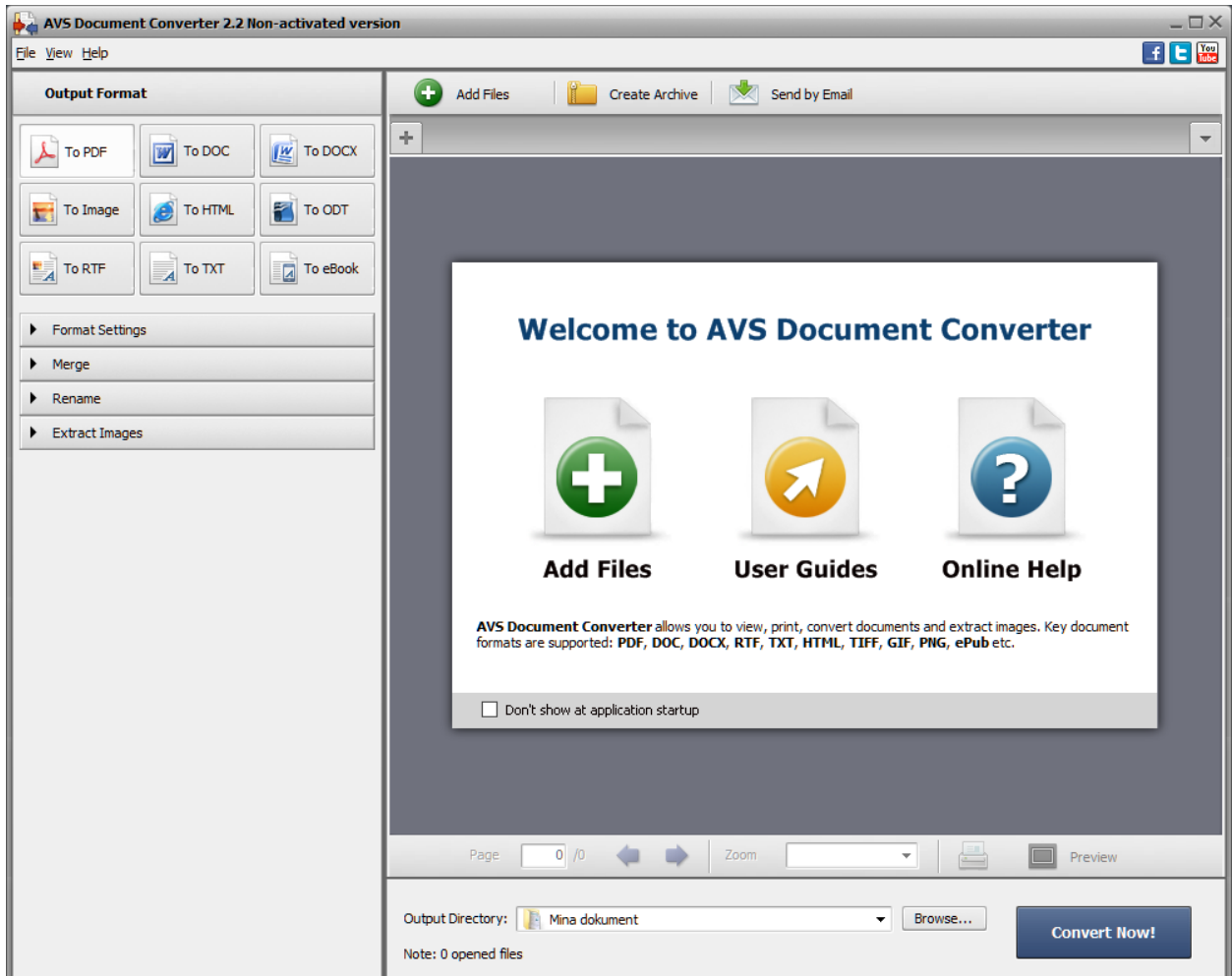


Figure 2: Start page

The User Guides are on-line guides for *all* the AVS tools. The Online Help contains instruction videos for the document converter.

2.4.1 Adding source files

The command “Add files” let you choose source files to work with. You can add many files simultaneously.

Loading TIFF image files

When I tried to add the seven TIFF images simultaneously, the program stopped working.

When I added only one TIFF file at a time, it still was impossible to load the largest images (397 MB) and the program stopped working, but the second largest (150 MB) was successfully loaded. It was also possible to add the five smallest files (154 KB up to 29 MB) simultaneously without problems.

Loading DjVu image files

I tried to add the 106 DjVu files simultaneously. The following message was shown:

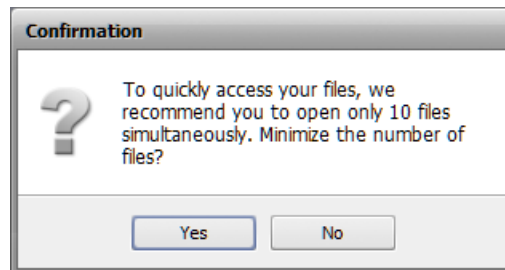


Figure 3: Warning message

However, adding the 106 files simultaneously went rather fast. I tested this several times, and occasionally the program stopped working, but most times the loading was successful.

2.4.2 Converting files

There are nine main choices of target formats (PDF, DOCX, Image, HTML, and so on). After selecting Image, you choose one of the following: JPEG, TIFF, GIF, PNG. For JPEG, you can also choose image quality of output files (default value is **100%**). For TIFF, you can choose **No Compression** or **Lzw**.

When I added TIFF files as source files, selected JPEG as output format, and then clicked "Convert now", the following message was shown:

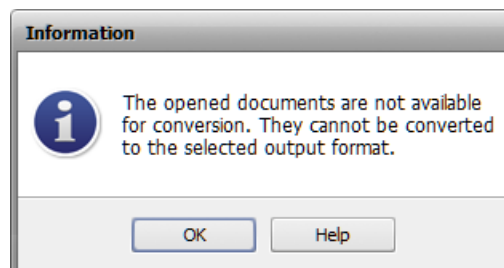


Figure 4: Error message

The same happened with the other image output formats. Thus, it is not possible to convert TIFF to other image formats with this tool. Neither can you use the tool to compress TIFF files.

Converting TIFF files to PDF

When PDF was chosen as output format, the conversion *seemed* worked for five of the loaded TIFF images. However, for two of the images, the result image was only a grey area. These documents were scanned pages of church books (the sizes of the source files were 154 KB and 29,1 MB, respectively), and the size of the target files was very small (59 KB).

The second largest (150 MB) could not be converted at all (the program stopped working).

For the remaining three TIFF images, the conversion worked and the quality of the output files was good. The sizes of the files were reduced:

- 5,61 MB → 1,34 MB
- 1,54 MB → 908 KB
- 13,4 MB → 2,4 MB

Converting DjVu files to PDF

Simultaneous conversion of 31 of the DjVu images went fast. All files were successfully converted, and the quality of the output was good. However, the target PDF files were much larger than the source DjVu files (approximately, ten times larger).

When converting 21 other DjVu files simultaneously, the program stopped working after converting nine of them. Furthermore, all the target files was of size 60 KB, and only contained white pages images.

A difference with the successfully converted 31 files was that these 21 files were larger; the *largest* file in the batch of 31 files was 336 KB, while the *smallest* file in the batch of 21 files was 1272 KB. When this large file was converted alone, it only seemed to have succeeded; the conversion resulted in the same white page as before. An attempt to convert the largest file (5200 KB) made the program stop working.

The rest of the files followed the same pattern: large files were impossible to convert, and some files were only seemingly converted.

Converting DjVu files to JPEG

Simultaneous conversion of 31 of the DjVu images into JPEG went fast. The chosen scale factor was 100%. All files were successfully converted, and the quality of the output was mostly acceptable (however, the esthetic value of some pencil sketches were much diminished). The target JPEG files were larger than the source DjVu files but moderately so (approximately, twice as large).

The previously unconvertable 21 other DjVu files were now successfully converted, and the quality of the output was good.

The rest of the files were converted without problems.

Finally, an attempt was made to convert all 106 DjVu image file simultaneously. This made the program stop working (after conversion of ~35 files).

Converting DjVu files to PNG

As for JPEG, the simultaneous conversion of 31 of the DjVu images to PNG went fast, and the rest of the files could also be converted (note that no scale factor can be selected for PNG). All files were successfully converted, and the quality of the output was acceptable. The target PNG files were larger than the source DjVu files (approximately, four to five times larger).

Again, the attempt to convert all 106 DjVu image file simultaneously made the program stop working (after conversion of ~45 files).

2.4.3 Recommendation

This tool should only be used when you want to convert small amounts of files. It is not reliable for batch conversion. Furthermore, check the result if you convert to files to PDF (the conversion sometimes results only in a white or grey page instead of an image). You may also want to check the quality of the images if you convert to JPEG or PNG (the quality varied between barely acceptable and good).

Note that the PDF files that are generated with this tool are not necessarily in the specific **PDF/A-1** format. Even so, they may conform to the rules of PDF/A-1, but this must be tested (see <http://www.pdfa.org/2011/08/validating-pdfa/>). It is possible to convert PDF to PDF/A-1, but not necessarily easy.

Grade

On a scale from 1 (very bad) to 5 (very good). X is "Not applicable".

Simplicity of installation: 4

Simplicity of management: X

Ease of use: 3 (would be higher if pop-up window messages didn't appear all the time)

Generality of solution: 4 – 5

Quality of result: 1 – 2