



PDFspy

version 6

Description

PDFspy is the ultimate “get info” utility for your PDF documents. It can extract a comprehensive list of attributes from a PDF file attach to a Switch job as an XML dataset.

NOTE – This app requires the separate purchase of Apago’s PDFspy software. More information can be found at <https://apagoinc.com/pdfspy>

New features and enhancements including:

- Support for PDF 1.7/ISO 32000 (Acrobat 9, X, DC)
- Element now shows CMYK separations that are actually used by text and vector elements
- New element that shows the number of shading objects in PDF file
- Restored output being written to stdout if -o option not used, recommend using -quiet option when writing to stdout
- Fixed calculation of page labels
- Improved text extraction algorithm
- Calculates color simulation values for ICCBased, Separation and DeviceN colorspaces
- Improved Unicode, ISO Latin and AdobePDF character set support

Some of the many types of information PDFspy can extract:

- Page information (count, size, boxes)
- Fonts usage (name, type, embedding & subset status, use of Unicode)
- Colorspaces used (alternates, separation names, index bases)
- Images (size, resolution, compression, colorspace)
- Use of transparency, smooth shadings and patterns
- Presence (or absence) of hidden text and optional content/layers
- Hyperlinks (size, location and destination)
- Annotations (size, location, type, contents, colors)
- PDF/X compliance (including output intent details)
- Metadata (info dictionary & XMP)
- Security and Encryption settings

Example uses:

- Asset management system: extract page count, metadata, font & image information
- Document management: determine text or image only documents, extract comments
- Preflight: extract information about colorspaces, compression & font types
- Developers: easily examine the structure of complex PDF documents

Compatibility

Switch Fall 2022 and later

The PDFspy app requires at least one input and one output connection. The output connection is a traffic light type. If the task succeeds, the job will be sent to the "Success" connections. Otherwise, it will be sent to the "Error" connections.

Flow elements properties

- Dataset – name for the XML dataset
- Fonts – include/exclude for font usage for each page and entire document including name, usage, type, subsetted, encoding, etc.
- Color spaces – include/exclude for color spaces for each page and entire document including base and alternate colorspace, DeviceN simulation values, etc.
- Images – include/exclude for images including colorspace, bit-depth, compression, location, pixel and physical dimensions and effective resolution
- Patterns
- Shadings
- Transparency
- PDF/X and PDF/VT information
- XMP packet
- Layers – include/exclude for optional content groups (OCGs) including name and properties
- Bookmarks – include/exclude for outlines
- Annotations
- Links
- Actions
- Names
- Form fields
- OPI comments
- Text – extract text from PDF pages
 1. Words – collect individual characters/glyphs into words. Includes font information and bounding box
 2. Columns – additional processing to collect words into nested collection of lines, paragraphs and columns
- Ink usage – calculate ink usage for each page
 1. Coverage size (mm) – sets the area used to calculate average ink usage
- Pages to process – selects which pages to process - either All or a custom range
- Limit to page area – selects the page area to collect information from – media, trim or bleed. Only objects with-in the select page area will be included in XML dataset.

Outgoing connections properties

- none

Support information

If you have any questions on how to use PDFspy, please send an email to support@apago.com or call +1 770-619-1884.

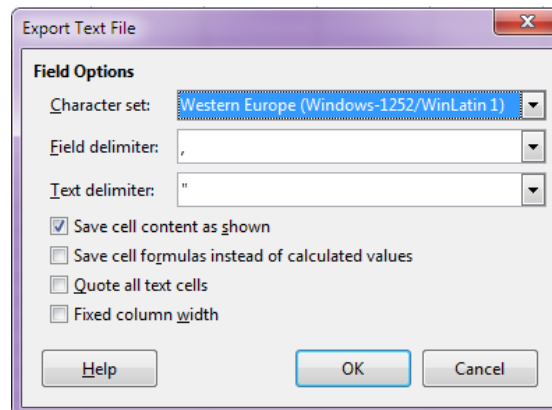
Tutorial

Goal: Select an imposition template using PDF file's trim size and page count

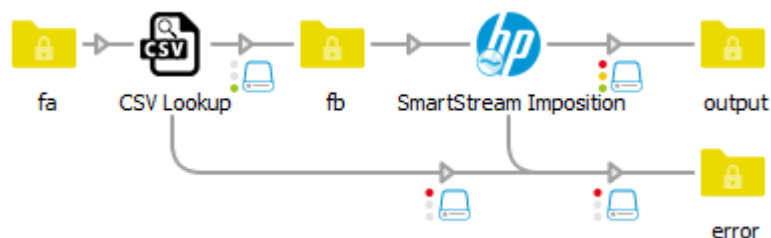
First we need to create a simple CSV that contains columns for the page width & height, page count and the imposition template name. You can use a spreadsheet program such as Excel or LibreCalc or just use a text editor. Make sure you include the names of each column in the first row.

	A	B	C	D
1	Width	Height	Count	Template
2	216	360	1	1sided_bizcards_h
3	360	216	1	1sided_bizcards_v
4	216	360	2	2sided_bizcards_h
5	360	216	2	2sided_bizcards_v
6	612	792	1	letter

Export the spreadsheet to CSV format. Use commas, semi-colons or tabs to delimit fields. Values that contain the delimiter character should be quoted. Here are the recommended settings:



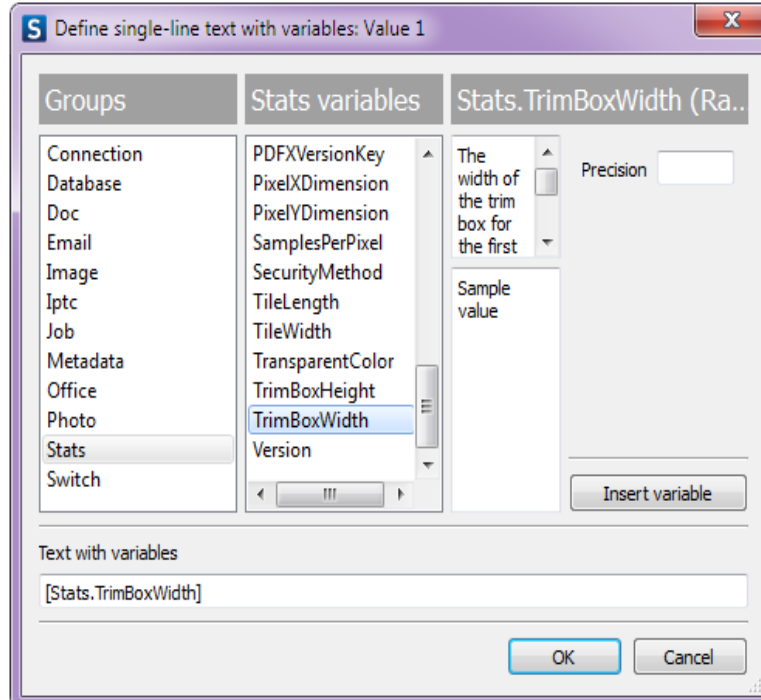
Next, create a flow containing CSV Lookup and your imposition program.



In the properties for CSV Lookup select the CSV file you want to use. Enter the names of the "Width", "Height" and "Count" columns for the "Compare 1", "Compare 2" and "Compare 3" properties and "Template" for "Lookup 1 column" property.

Property	Value
Name	CSV Lookup
Description	
CSV file	\\Data\templates.csv
Column delimiter	Automatic
Compare 1 column	Width
Value 1	
Compare 2 column	Height
Value 2	
Compare 3 column	Count
Value 3	
Lookup 1 column	Template
Private data key 1	
Lookup 2 column	None
Lookup 3 column	None
Lookup 4 column	None

Now we have to select where Switch should get the values that will be used for each comparison. In this example we will use Switch variables to retrieve the trim box width height and the number of pages from a PDF. For "Value 1" properties, which will be compared against the values in the "Width" column of the CSV, we select "Define single-line text with variables" and then "Stats" and "TrimBoxWidth".



Use the same steps to select the variables for "Value 2", [Stats.TrimBoxHeight] and "Value 3", [Stats.NumberOfPages].

CSV Lookup uses a feature of Switch called "private data" to store the values it retrieved from the CSV file. Private data is stored in a job's metadata and can be easily used later in the flow using Switch variables. Private data has a key, which is used to name it, and a value. For this example, set the "Private data key 1" to "template_name".

Compare 1 column	Width
Value 1	Single-line text with variab...
Compare 2 column	Height
Value 2	Single-line text with variab...
Compare 3 column	Count
Value 3	Single-line text with variab...
Lookup 1 column	Template
Private data key 1	template_name
Lookup 2 column	None
Lookup 3 column	None
Lookup 4 column	None

The final task is to change the imposition program's properties so it retrieves the name of the template. Select "Single-line text with variables" for "Imposition Template" property and then Job group and PrivateData variable. Enter "template_name" for the key.

