

Paper 72-25

Creating Adobe® Acrobat™ (PDF) Documents Directly from SAS® Software

Paul Wehr, Industrial Softworks, Ann Arbor, MI

ABSTRACT

The Adobe® Portable Document Format (PDF) specification allows complex documents to be shared across platforms and operating systems, independent of the software used to develop them. For SAS output, PDF can be used to conveniently distribute reports to a variety of clients, store the SAS component of a much larger document, such as a pharmaceutical new drug application, or archive historical results. Although tools such as the PDF virtual printer driver, and the Adobe Distiller allow anyone to create PDF files, it does require additional software, and a separate, manual step. The SAS2PDF macro was developed to provide the ability to convert standard SAS output to the PDF specification within the SAS job, and without the need for any additional software.

ABOUT PDF

While PDF has been overshadowed somewhat by HTML as an information exchange format, PDF has some notable advantages over HTML. PDF provides much greater flexibility in creating complex documents, containing diagrams and images. PDF is also better suited to managing large documents (100-100,000 pages) because of its indexed object format. Any single page in a document can be viewed almost as fast as the host system can render it. In addition to these advantages, PDF also shares most of the same platform-independence and hypertext features that HTML provides.

Why Use PDF for SAS output?

While SAS does provide some functionality for printing text on a page, it is either not specific enough to fully instruct an arbitrary printer (LINESIZE and PAGESIZE options), or the output is too specific (FORMS, which are targeted to a single, specific printer). A popular approach to providing text reports in universal way is to "import" the SAS text output into a word processor and set the font and margins so that the report would print as expected. While this manual approach may be appropriate for a small number of reports, it is not particularly practical in a batch environment where hundreds or even thousands of reports need to be generated for a wide variety of target audiences or on a regular basis. Furthermore, sending reports in a proprietary word-processing format requires that the sender and recipient have compatible software installed. Publishing reports as PDF documents through an automated process provides SAS-generated reports in a format that is convenient to produce while supporting high degree of generality, given the range of platforms for which the Acrobat Reader is available.

TECHNICAL DETAILS**PDF Structure**

The organization of a PDF is designed to follow a tree, or outline structure, where each element of the document, including page size and orientation, fonts, images, and text are branches of the tree. The speed and efficiency of the PDF file is realized by indexing the starting position of each of these elements. This index can occur anywhere in the document, but is placed at the end of the file by the SAS2PDF macro, because that is the point at which the start positions of each object are actually known.

Implementation:

The approach to casting standard SAS output to a PDF file in

this macro is relatively straight-forward. The macro first provides a series of basically static header information, setting up the paper size and orientation, selecting an appropriate [fixed-width] font: Courier. After the setup is finished, it simply copies each line of output in the ASCII source file to an equivalent line on a page in the PDF file.

The only delicate task in creating the PDF file is being able to track the exact byte-wise starting position of each object in the file for the index at the end. In the SAS2PDF macro, this is handled by the OUTPUT: block. The OUTPUT block acts as a gate-keeper, which counts and tracks the length of every string as it is written to the output file. The OUTPUT block stores these results in an array, which gets printed at the end of the file, in the cross reference table.

OTHER ALTERNATIVES

The times they are a-changin'. Since I submitted this paper, I've learned that there are plans for ODS to support PDF as a native output driver with release 8.1. Since ODS is clearly the right way to do output from SAS procedures, it would effectively replace this macro. For V8 users, there is also an experimental ODSPRINTER driver (<http://www.sas.com/rnd/base/topics/odsprinter/index.html>) which can be linked to the Acrobat printer driver to automate the process of generating PDF files.

There are also some tools similar to SAS2PDF available. One is a shareware program, called "TXT2PDF," written in Perl (<http://www.sanface.com/>). This program reads a plain text file (like SAS2PDF) and converts it to PDF, although it provides a few more options than SAS2PDF. Another tool is HTMLDOC (<http://www.easysw.com/htmldoc/>), which expects HTML as input, but can generate an indexed PDF as output.

Finally, if you are serious about PDF publication, (or document publication in general) and need a lot more control than any of these tools provide, a free, and open-source document publication environment called LATEX (<http://www.tug.org/>) is really worth investigating. If you pursue this route, ODS will support you with A LATEX ODS driver is planned for version 8 (<http://www.sas.com/rnd/base/topics/expv8/>).

REFERENCES

"Portable Document Format Reference Manual" - Adobe Systems Incorporated, 1993

CONTACT INFORMATION

Paul Wehr
Industrial Softworks
2181 Stonebridge Drive South
Ann Arbor, MI 48108
(734) 944-7900
paul@industrialsoftworks.com

The code and documentation for the macro described in this paper can be obtained from the Industrial Softworks web site at <http://industrialsoftworks.com/>