# COMPARISON OF FILE FORMATS FOR EXCHANGE OF FULL PAGE ELECTRONIC DATA

Prof. Dr. Kurt Schläpfer, Erwin Widmer*

**Keywords:** Prepress, Files, Electronic Data

**Abstract:** The present paper gives a comparison of file formats which may be used for exchange of digital files which represent completed or multiple pages. The following formats are explicitly addressed: TIFF, TIFF/IT, EPS, DCS and PDF.

To compare the file formats, digital test documents comprising different elements were produced, converted into the different file formats and tested on different hardware and software RIPs.

The comparison addresses the following aspects of file formats:

- workflow to convert a file from an application program to the exchange format
- file editability
- legal usage of type fonts
- file size
- application of imposition programs
- usage of data compression
- rendering of traps, spot colors, vignettes etc.
- availability of preview images
- compatibility with color management.

---

* EMPA/UGRA St. Gallen, Switzerland

## Problem

With the increasing significance of computer-to-plate and digital printing systems, it is becoming a real necessity for prepress companies and customers to supply printing companies with digital data in the form of full page electronic data rather than with films or pure text data.

The simpliest possibility is to exchange data in the form in which it is created, i.e. in the «native» application program, e.g. Quark XPress. Such a data file (referred to in the following as native data file), however, has a number of serious disadvantages for the data exchange:

- The data recipient may not have the same version of the program or the same program modules (extensions, plug-ins) as the data supplier.
- The data recipient may not have all the type fonts that the data supplier has used in the data file. The exchange of type fonts, however, is illegal.
- The native data file could be changed or edited by the data recipient, what is not desired.
- Data files, which are imported into the application program, are usually attached and not embedded. In such cases, there is a danger that they will be not found, because the path has been changed.
- Image files cannot be exchanged in a compressed form.

For the exchange of data, it therefore follows that the native data file has to be converted into a file format, not exhibiting the aforementioned disadvantages.

In general, data exchange calls for a file format that meets the following requirements:

- Type face embedment is possible
- Data are not editable
- Data can be displayed on screen
- Size of the exchange file is not much larger than that of the native file
- Bleedoff elements are retained
- Traps are retained
- Data compression is possible
- File format can be generated directly from the most common application programs

- File format legible with PostScript RIPs (levels 1 and 2)
- Images can be described with device-independent color data.

In the following chapter, the characteristics of some data formats which may be considered as exchange formats are described.

## Characteristics of some data formats

### PS (PostScript)

PostScript (PS) is a page description program which is especially suitable for object-oriented descriptions (description of text characters and geometric elements). Since the coding is made with more than 200 operators, data files consisting of different elements, such as graphics, text, etc., are complex and error-prone. As the numerous PS generating drivers of various origin, especially the unlicensed PS drivers (clones), do not always produce identical code, they are interpreted differently on the various PS RIPs.

One disadvantage is that PS data files are organized sequentially, i.e. optional access to a single page is not possible.

Moreover, the RIP time of a PS data file depends greatly on the type of image and graphic elements it contains.

### EPS (Encapsulated PostScript)

EPS was derived from PS in order to import PS pages in page make-up programs, such as Quark XPress. For this purpose, EPS files contain not only the PS commands, but also additional instructions which include information about the data file (e.g. size) to be imported and enable a preview on the screen.

Based on the concept that EPS files are created for embedment in a document, they always describe a single page the size of which has to be defined by a bounding box. Certain PS operators are forbidden for EPS data files. If this fact is ignored, the EPS data file will be illegible on PS-RIPs. In principle, the EPS data files are not editable. However, certain application programs, e.g. Adobe Illustrator, generate EPS data files that are still editable.

A disadvantage is that traps can be lost as a result of conversion into an EPS data file, e.g. when Quark XPress is the native file.

### DCS (Desktop Color Separation)

DCS is a further extension of the EPS format. It was developed to make the separation data of an image addressable as single color files. The master file is a normal EPS data file, which is used as a composite file when an output is made on a color printer which needs no separate files. On output devices requiring single color files, the output of DCS files is faster, because the RIP processing may start after the first color file is transferred. This is favorable when storage capacity is limited, or if the data transfer to the RIP is effected very slowly. A DSC file also features a preview image.
In contrast to EPS, traps remain in the DCS format.

### TIFF (Tagged Image File Format)

This file format was designed for cases where the data input is effected via a scanner. The data are therefore coded as a grid of pixels, which is efficient for images, but not necessarily for text, graphics and other elements that can be defined by vectors. TIFF files cannot be produced directly from application programs (e.g. like Quark XPress), and therefore require the aid of a special program. The output parameters such as resolution, format etc. are defined by tags, whereby not enough tags are defined for the requirements of highend image processing.

### TIFF/IT (Tagged Image File Format/Imaging Technology)

This file format is a further development of TIFF for the specific requirements in highend image processing. It was developed on the basis of the former Digital Data Exchange Standard (DDES defined for magnetic tape as data carrier) and is now awaiting approval as an ISO Standard (ISO 12639).

In general, a TIFF/IT data file consists of three types of data files:

- Color continuous-tone image data (CT): This data is used to code images pixel-wise.

- Line art image data (LW): This file type comprises graphic elements and text. The resolution of this file is appreciably higher than the CT file (usually more than 350 dots/cm).

- High resolution continuous-tone image data (HC): This part describes high resolution edges to continuous-tone images. HC data are only present where this is necessary from the image content.

These files are combined in the Final page file format (FP).

Due to the raster structure, the RIP time is, in contrast to PostScript files, only dependent on the resolution and print format, but not on the type of objects (text, linework, etc.). In addition to CMYK data, TIFF/IT allows to define RGB and CIELUV data. A disadvantage is that the preparation of TIFF/IT files from application programs such as Quark XPress necessitates special programs. Moreover, the TIFF/IT format is not widely used in Europe.

### PDF (Portable Document Format)

PDF was created by Adobe claiming to provide a universally utilizable file format. Although it emerges from PostScript, it eliminates several disadvantages, such as its complexity, and offers new possibilities, e.g. hypertext applications. PDF is not a programming language (like PS) and is therefore more efficient and less data intensive. Type fonts that are not provided in a PDF data file are generated by means of font substitution.

PDF files are editable in some cases. With Acrobat Exchange it is possible to correct lines of text and to exchange or add complete pages. With the Adobe Illustrator (from version 5.5) and Macromedia Freehand (from version 7), it is possible to edit all PDF elements.

Legal type face embedment is possible, because the file only contains the characters that actually occur in the document. If the number of characters used in the document exceeds a certain percentage (default setting = 35%), the complete font and not the used subset of characters is embedded in the PDF file. The fonts, however cannot be extracted from the PDF data file and used further. Therefore, the majority of font manufacturers permit the embedment of their fonts in PDF documents.

PDF files also permit direct access to any desired page of the document.

The preparation of PDF files starts from PS data. PS errors are therefore recognized before they reach the RIP of an output device. The conversion into a PDF data file may therefore be considered as a «pre-ripping» test.

As PDF files are converted back into PS files prior to output, this leads to a standardized PS code which ensures maximum readability on PS RIPs.

Whereas PDF provides a general description of the data format (at present with version 1.2), Adobe has created the relative software with the name Acrobat. The latest version 3.0 is available for various operating systems.

The following components of the Acrobat software are important for the exchange of data:

- Acrobat Reader: required for reading a PDF file on the screen and for the data output. This component is available free of charge.

- Acrobat Exchange: includes the functions of Acrobat Reader and certain editing possibilities for PDF files, for example, the
  - insertion of pages
  - addition of notes
  - changing of page sequence
  - correction of individual text lines.

- As the file recipient only needs the Reader module to output the data, this module may be provided together with the exchange file. Alternatively, Acrobat Reader may be downloaded from the Adobe homepage (www.adobe.com).

- Acrobat Distiller: used to produce a PDF file from PS data.

In all probability, PDF files are expected to be created directly from application programs in the near future.

## File size resulting with different data formats

An important criterion for the practical performance of a data format is the resulting file size when data are converted into this format. Clearly, the file size depends much more on the type of elements forming the content of the file. Whereas, for instance, type matter contributes little to the file size, images may lead to very voluminous files.

To test the file size resulting with different data formats, 4 test documents consisting of one DIN A4 page were created:

| | |
|---|---|
| Doc # 1: | Monochrome image (raster file) |
| Doc # 2: | 4-color image (same picture as Doc # 1) |
| Doc # 3: | Graphics (vector file) |
| Doc # 4: | Typical catalog page consisting of text, color backgrounds and images. |

The test documents were created with the page make-up program Quark XPress on a Macintosh workstation. The steps to create the documents and to convert the native file into the different data formats is shown in figure 1.

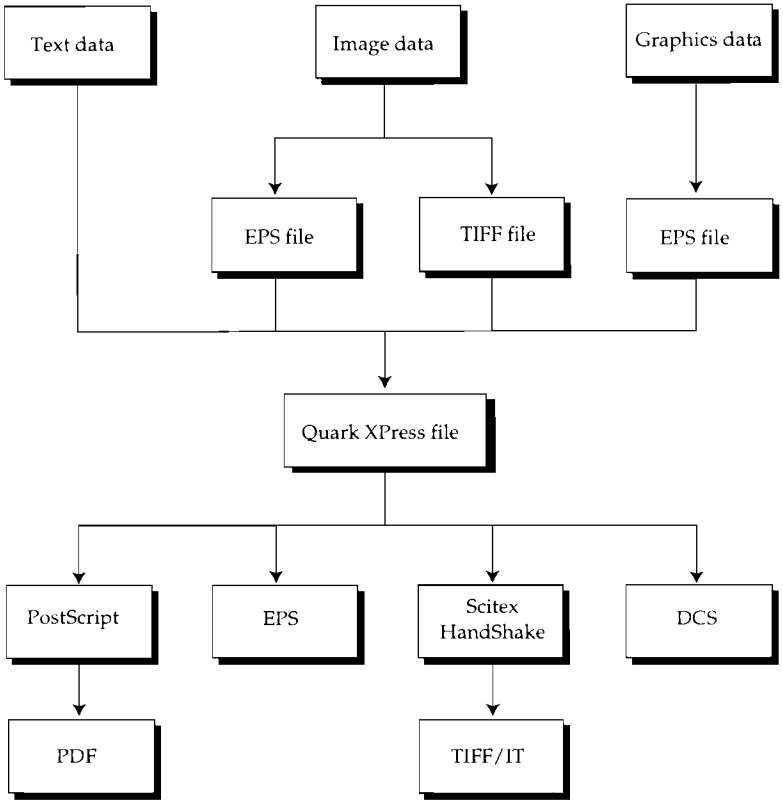


Figure 1 **Workflow to create the test files and convert them into the exchange formats**

The resulting file sizes are listed in table 1. As can be seen, the conversion to EPS and PDF files leads to file sizes which are comparable with the native file, and this independent of the type of document. In the case of PostScript and DCS files, however, the file size depends largely on how the original image file is imported into the application program. If a 4-color image is imported as EPS file, the file size becomes 3 to 5 times larger than when imported as TIFF file.
A very data-intensive file format is DCS. Here the resulting file sizes can be up to 5 times larger than the native file.
A very «lean» file size is obtained when documents are compressed in the PDF format. The final document size is then typically not more than 30 % of the native file size.

The TIFF/IT format which has not been tested with data compression leads to file sizes which are considerably larger than PDF files. Based on this fact, PDF is the more attractive exchange format than TIFF/IT.

| Test document | File size (MB) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Native file * (QXP) | Post-Script | EPS | DCS | PDF | PDF compressed | TIFF/IT ** |
| **Doc # 1** Monochrome image | | | | | | | |
| – imported as EPS | 4.7 | 4.7 | 4.5 | – | 4.1 | 0.4 | 11 |
| – imported as TIFF | 4.8 | 4.2 | 4.4 | – | 4.0 | 0.4 | 11 |
| **Doc # 2** 4-color image | | | | | | | |
| – imported as EPS | 21.6 | 85.3 | 17.2 | 106.7 | 16.6 | 3.1 | 36.9 |
| – imported as TIFF | 17.5 | 15.9 | 16.6 | 32.2 | 16.2 | 3.0 | 36.9 |
| **Doc # 3** Graphics (imported as EPS) | 2.2 | 6.6 | 1.7 | 7.6 | 2.8 | 0.9 | 43.6 |
| **Doc # 4** Catalog page (images imported as EPS) | 13.0 | 60.0 | 12.9 | 72.9 | 16.0 | 3.8 | 36.9 |

* including fonts and images
** CT = 120 l/cm, LW = 720 l/cm

Table 1 **File size resulting with different data formats**

## Is editability necessary for data exchange?

As various organizations have already concerned themselves with the issue of data exchange, different opinions exist concerning the need for editability of exchange files.
The first thought is that editability is not desired, unless it is under control of the file creator. A second consideration, however, reveals that some degree of editability may be desired at the receiving site, for instance, for exchanging entire pages, adding new pages, or changing page numbers. Therefore, depending on the particular case, the answer to the question put in the title is «yes» or «no». Under these circumstances, an ideal exchange format should offer two options, i.e. one for «read-only» and another for allowing limited changes. As will be shown in the next chapter, PDF is a format offering these features.

If an exchange file is sent to a printing company, it is an exception that the file can be used without any further manipulation. The minimum action necessary by the printer is to import a control element for dot gain and density measurements and – in case of a multi-page document – to make the page imposition.

As the currently available imposition programs are based on PostScript data, the requirement to an exchange format is that it can be converted to PostScript.

An issue which came into discussion only recently is whether traps should be defined by the file creator or at the receiving site. The ongoing discussion tends to see the printer as the right person to define traps.

## Comparison of data formats

Table 2 shows a comparison of data formats which can be considered as suitable for exchange formats. The extent to which a data format will fulfill the requirements with respect to data exchange depends chiefly on the application program that is employed for the conversion of the data. The comparisons made in table 2 relate to Quark XPress as original data file.

On the basis of the assessed characteristics in table 2, PDF (Portable Document Format) is recommended as file format for data exchange. The Adobe Acrobat 3.0 program is required for its application.

The workflow with PDF is shown in figures 2 and 3.

| Requirements | Requirements fulfilled with | | | | |
|---|---|---|---|---|---|
| | PS | EPS | DCS | TIFF/IT | PDF |
| Type face embedment possible | yes | (yes)[1] | no | (yes)[2] | yes |
| Data not editable | (yes)[3] | yes | yes | (yes)[4] | (yes)[5] |
| Data displayable on screen | (no)[6] | (yes)[7] | (yes)[7] | yes | yes |
| File size not much larger than native file | no | yes | no | no | yes |
| Traps retained | yes | no | yes | yes | yes |
| Bleedoff elements retained | yes | no | no | yes | yes |
| Data compression possible [8] | no | no | no | yes | yes |
| File format can be generated from customary application programs | yes | yes | yes | no | (no)[9] |
| File format readable with PostScript-RIPs or convertible to PostScript | yes | yes | yes | (yes)[10] | yes |
| Images can be described with color data | yes | yes | no | yes | yes |

[1] Not possible with every application program; special software (e.g. FontIncluder) required.
[2] No font embedment, but conversion into a bitmap
[3] Only editable with special software, e.g. OneVision.
[4] Images editable
[5] Not editable with Acrobat Reader, partly editable with Acrobat Exchange
[6] Only possible with Display PostScript or GhostScript.
[7] Only possible if stored as bitmap preview
[8] In this case, it is not a question of whether data-compressed images can be imported in a document, but whether the exchange file can be compressed.
[9] Can only be generated from PageMaker
[10] Only possible with RIPs from just a few manufacturers

Table 2 **Comparison of data formats**

The decisive advantages of PDF for the exchange of data are:

- It facilitates a legal form of type embedment
- The data file is only editable under certain conditions
- Direct access is possible to individual pages of the data file
- The elements of a PDF data file (text, linework, images) can be integrated in a PDF form and transferred together with the exchange data.
- The job description can be integrated in a PDF form and transferred together with the exchange data.
- PDF files can be viewed on monitors using different platforms such as Macintosh, Windows and Unix. This allows the data recipient to check the file.
- The data supplier can detect and eliminate any existing PostScript errors during the preparation of a PDF file. The data recipient therefore acquires a perfect data file when PDF is converted to PS. In this sense, the preparation of a PDF data file is similar to a RIP operation, meaning that the data file can be considered as being «pre-ripped» when transferred to the recipient.

It should be noted, however, that PDF has also some disadvantages with respect to data exchange:

- PDF files cannot be imported into application programs such as Quark XPress, though an option exists to generate EPS files from PDF files. Tests have shown that this function does not properly work with complex documents.
- PDF does not fully support color management applications. For instance, the attachment of an input or output profile is not possible.

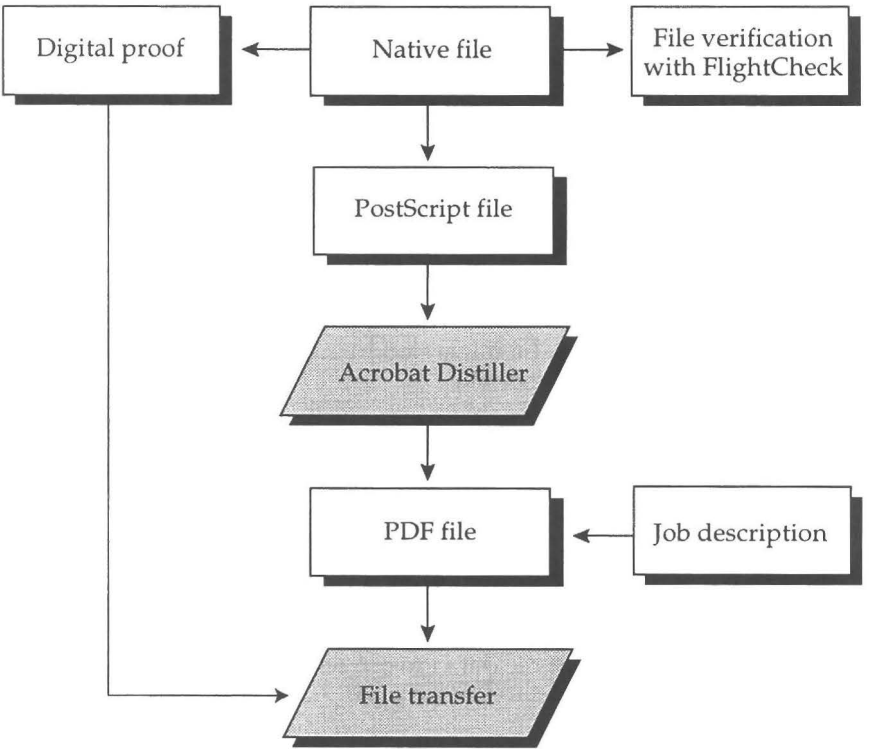


Figure 2 **Workflow for data supplier when working with PDF**

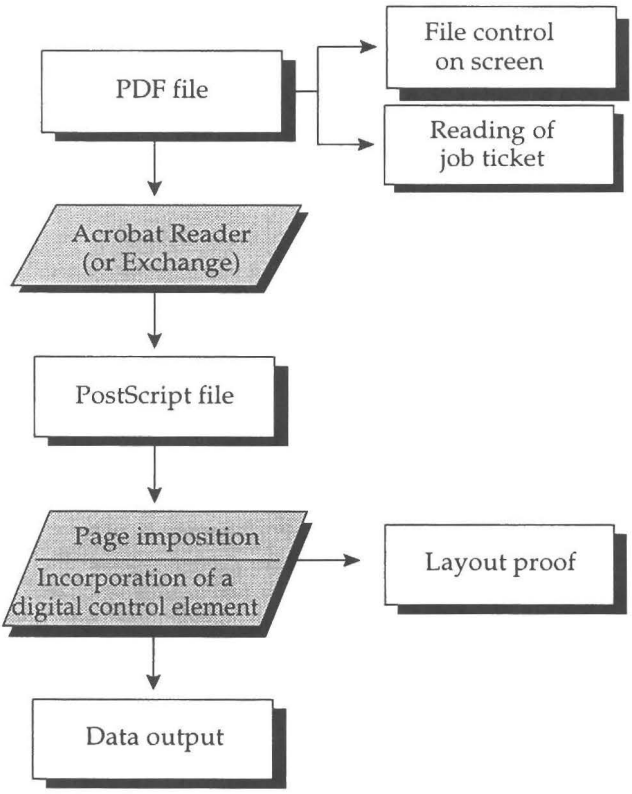


Figure 3 **Workflow for data recipient when working with PDF**