Specification for Prepress Digital Data Exchange

Prof. Dr. K. Schläpfer, Erwin Widmer*

Keywords: Prepress, Digital, Data , Exchange, PDF

Abstract: Various organisations have already concerned themselves with the subject of data exchange and made corresponding recommendations. Whereas some of these recommandations only relate to the exchange of advertisements in periodicals, others are more comprehensive in their range and application, but do not offer specific working procedures.

Ugra (Association for the Promotion of Research in the Graphic Arts Industry in Switzerland) has decided to prepare specific guidelines for the handling of data exchange. To ensure that they can be applied in the practice, investigations have been carried out with a test document and on various output devices.

The Ugra specification is based on the PDF file format. The specification explains the workflow from the data provider to the receiver.

The starting point is the «native» file. The most important steps are:

- preflight checking
- proofing
- conversion to the exchange format
- definition of the job ticket
- data transmission
- trapping definition
- page imposition

---

*EMPA St. Gallen, Switzerland

Introduction

This paper provides guidelines for the exchange of digital data between a data supplier (which can be a prepress company or a customer) and a data recipient (which is usually a printer).

The purpose of these recommendations is to enable the data supplier to produce a data file which can be processed in a perfect manner without having to contact the data recipient. Furthermore, these recommendations shall ensure that any problems that may arise can be identified and solved without any recriminations.

Scope

These recommendations are valid for the exchange of data files which are «ready for press». The term «ready for press» is defined as follows:

• The data file is formatted in the form of full pages.
• The data file can be processed correctly on the RIP of the respective output device without any corrections or modifications.
• The image data are either CMYK data suited for the output device or the printing process, RGB data supplied with a colour profile for the employed input device (e.g. scanner), or CIELAB data.
• The data file must permit an imposition of the pages by the printer.
• The printer must be able to incorporate digital control elements according to his requirements.

The recommendations made here do not apply to the transfer of «raw» data (text, graphic, images) and where the printer has to process or correct the data file prior to printing.

The data file(s) provided by the data supplier should be suitable for the printer even if he does not have any devices or equipment that are available to the data supplier.

The recommendations are valid for the production of printed matter for which the data can be produced according to the workflow shown in figure 1. These are:

> • Advertisements for periodicals
> • Catalogues
> • Commercial printing
> • Brochures

- Annual reports
- Books
- Manuals

These recommendations also cover printed matter which do not consist of pages in the conventional sense, such as posters and forms.

The recommendations, however, do not apply to the production of periodicals, because they are produced according to another workflow than shown in figure 1.

Production information for the data file to be printed

If a data file is transferred to a data recipient, it may be assumed that a job description has already been made and forwarded to the data recipient. Consequently, the data file has only to be accompanied by an information describing the data file and the data output.

In particular, the following information has to be provided:

Description of the submitted data

- File format
- File elements (text, linework, images)
- Type of compression
- Employed type fonts
- UCR/GCR setup
- Image resolution
- Total file size in MB

Information for the data output

- Type of screen (FM or conventional)
- Screen width or dot size for FM screening
- If CMYK data: printing standard (including dot gain, solid-tone density, process colors)
- Definition of the spot colour(s) (if available)
- References to duotone images (if applicable)

The information describing the data file and the form of data output may be integrated in a so-called job ticket, accompanying the data file.
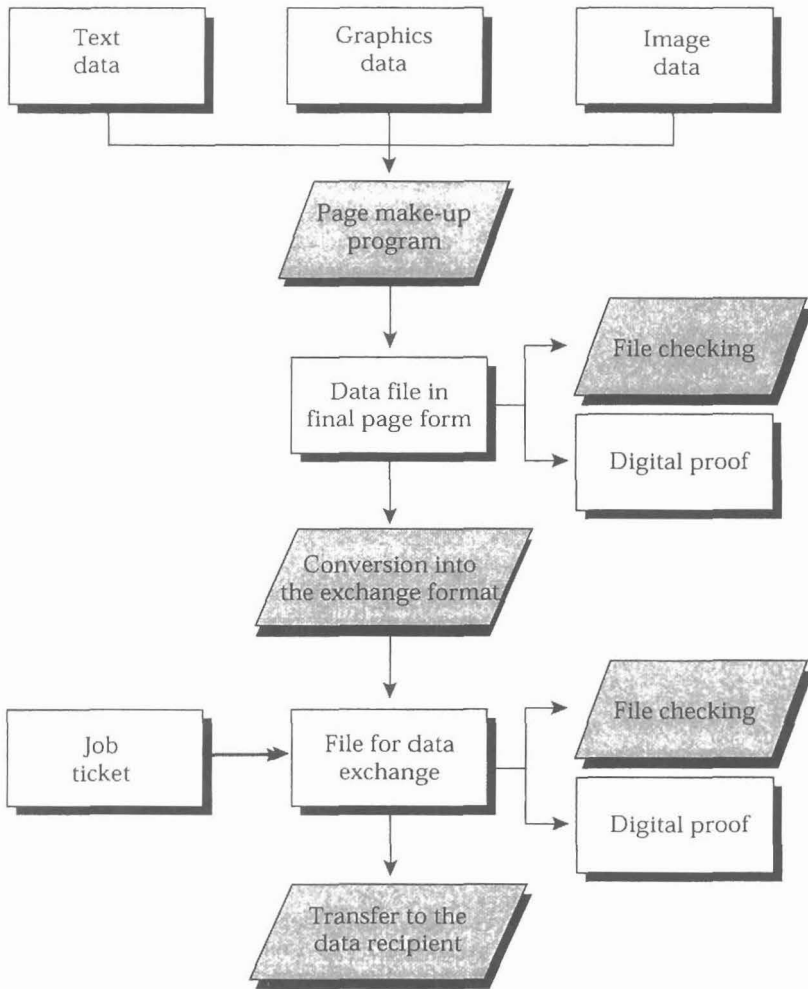
Figure 1: Workflow for the preparation of files supplied to printers

Furthermore, the data files are to be accompanied by a digital proof and a preflight checking document (see section «Control of data by the data supplier»).

Should the data file be transmitted by means of telecommunication, these enclosures are to be forwarded per separate post or by courier.

File format for the exchange of data

A comparison of the formats suitable for data exchange is made in the authors 1997 TAGA paper «Comparison of file formats for the exchange of full page electronic data».

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

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 compressed individually, which is important for the data exchange.
• 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 Apple-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. 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 that PDF files, for the time being, cannot be imported directly into QuarkXPress. For this reason, the PDF format must be transformed into an EPS file in order to be imported into a document with many pages.

Control of data by the data supplier

The biggest problem with data exchange are errors in the file, which either impair or make further processing impossible. In principle, a differentiation has to be made between two forms of error:

- Errors in the content: These include missing type fonts or image files, incorrect paths and wrong settings with regard to colors, resolution, page elements, etc.
- Errors resulting from file conversion: With PostScript, for example, different drivers may produce different commands which again are interpreted differently on RIPs.

PDF has been selected as the exchange format to eliminate the possibility of conversion errors (or to make them as unlikely as possible). Of all the exchange formats on the market, this data format is considered to be the least sensitive to conversion errors.

The errors in content can only be eliminated by subjecting the data file to a systematic and, if possible, automatic control. Programs known under the name «preflight checker» are suitable for this task. There are a series of such programs on the market, which differ in their range of performance and applicability for various data formats, application programs and platforms. For the present recommendations the «FlightCheck» program (manufactured by MarkzWare), was tested and found to be suitable for the described purpose.

FlightCheck reads native files such as QuarkXPress and Adobe PageMaker as well as PostScript files.

It verifies the completeness of the type fonts and images and the correctness of the image resolutions, color definitions and traps. It is possible to scan up to 140 characteristics.

The user can define a report issued by the program with the aid of its basic setting (GroundControl). The report provides information about the

- employed process inks and spot colours
- fonts used in the file
- names of all images with resolution, file size and original data format.

There are now preflight checkers on the market, which check PDF files. These tools helps the data supplier to check the PDF file before it is supplied and the data receiver to check the PDF data file when it arrives.

Proof preparation

It is the task of the data supplier to provide a proof.

In case of multi-color colored half-tone images, the proof – which obviously has to be established digitally – has to simulate the production print exactly (i.e. within the agreed tolerances) with regard to the following characteristics:

• color and density of the primary colors and any spot colors
• dot gain
• color of the printing substrate.

In the case of images a proof is always to be in full-scale (1:1). However, one cannot expect that the proof be made on the production stock or that the gloss be simulated.

To establish a color proof that matches the production print, the data supplier is obliged to ask the data recipient about the printing conditions and the printing substrate. Equally important is the question of whether negative or positive films will be made from the data.

Wherever possible standard conditions should be agreed between the data recipient and the data supplier. Emphasis should be placed on the new ISO Standard for offset printing (ISO 12647, Part 2), which differentiates between five types of paper.

The data supplier will then supply the recipient with CMYK data corresponding with the data used for the proof preparation. Another way of working results in cases where color management is employed.
To make a color proof measurable, a digital print control strip has to be printed. The Ugra/FOGRA Digital Print Control Strip is suitable for this purpose.

Compression of exchange files

To achieve a good compromise between file size and image quality, it is advisable to apply the compression possibilities as follows:

• Text and graphic data are to be compressed lossless, whereby algorithms such as LZW or ZIP do not differentiate to any significant extent.
• For image data, the lossy JPEG compression is recommended. The «medium» position should be selected as the compression setting.

- A reduction of the resolution (down-sampling or sub-sampling) is only to be made when the resolution (pixels per inch) is more than twice the value of the screen width (lines per inch).

With JPEG compression and when reducing the resolution, it should be noted that this is always associated with a loss of information. When image data are reduced irreversibly, this is a modification of the image content and may even infringe the copyright.

Recommendations for the preparation of the native data file

The question of whether a data file can be processed by the data recipient without difficulty depends chiefly on the manner in which the native data file was prepared. The following recommendations may help to prevent several of the most frequent difficulties experienced with the output of data.

- The data file should not be larger than is necessary for the content. For example, unnecessary high image resolutions are to be avoided. In particular, the scanning resolution should not be more than twice the screen width, and only 1.5 times with fine screens (above 200 L/inch). Furthermore, graphics should be in vector format rather than in pixel format.
- Only PostScript Type 1 fonts are to be used. TrueType and system typefaces can lead to errors by the data output.
- When working without color management, the image data should always be CMYK (if necessary supplemented with spot colors), which are based on the production printing conditions. RGB and CIELAB data would need the use of color management, for which in conjunction with the PDF file format no software is available today.
- To avoid protracted output times, the rotation and scaling of pictures should be made in the image-editing program (e.g. Photoshop) and not in the page make-up program. If rotations are made in the layout program, integer angles lead to shorter output times than angles with decimal values.
- To ensure that cut-out images are produced in a correct manner, they should be made with the PostScript clipping path and not by means of manual cut-out techniques.
- Complex graphics, which are produced with FreeHand or Illustrator, may be simplified by subdividing the paths.

Recommendations for data recipients

The perfect processing of the data files can be simplified by observing the following hints:

- It is advisable to use only original PostScript RIPs and not clone-RIPs.
- The job ticket transferred with the data file is to be carefully studied before the file is processed.
- A layout proof which permits a check for the completeness of the data file and the correct position of the pages should be made after the imposition of a multi-page document.
- Calibrated devices should only be used for the data output. The linearisation curve of the device is therefore to be checked periodically. The Ugra/FOGRA PostScript Control Wedge, for example, can be used as a tool for the adjustment and checking of the device calibration.
- If an output with stochastic screening is required, this is only possible on RIPs with corresponding software. (The stochastic screen is not incorporated in the supplied data file).
- A print control strip, which is imported as digital file to be printed after imposition, is required for the adjustment and monitoring of the ink feed rate. The Ugra/FOGRA Digital Print Control Strip is an appropriate element for this purpose. It is suitable for output on both film and plate, as well as for use on digital printing presses.
- Although the employment of a Level 2-RIP is not a pre-requisite when working with PDF data, it is nonetheless recommendable.

Transfer of data to the recipient

The data can be transmitted to the recipient either via ISDN or by means of a mail service or courier. A magneto-optical disk with a diameter of 5 1/4 inch (with a current storage capacity of 4.6 GB) is recommended for the exchange of the material data.

The minimum equipment requirements for transmission via ISDN are an ISDN user access and an ISDN card. This will ensure a data exchange rate of 64 kbit/s, respectively 8kB/s, theoretically allowing the transmission of 28.8 MB/hour.

Since an ISDN user access has two channels each with 64 kbit/s, it is possible, in principle, to link these two channels and thus attain a transmission rate of 128 kbit/s. At the moment, however, this is only possible with just a few ISDN cards together with special software, requiring that supplier and receiver have a compatible ISDN card.

An appreciably higher data transmission rate can be realised with an ISDN connection offering 30 channels each with 64 kbit/s. However, these channels can only be linked with special soft and hardware, for which a network is usually a pre-requisite.

Therefore, a data carrier is still recommended in cases where it takes more than thirty minutes to transmit the data file.

In principle, an on-line service or Internet may also be considered for data transmission. The transmission via Internet can be effected in two ways:

- The exchange file is transmitted to the recipient as an attachment to an e-mail.
- The exchange file is transmitted in the ftp mode (ftp = File Transfer Protocol). In this case, the data supplier saves the exchange file with a password on his server and informs the data recipient that he can «download» the data file by means of file transfer.

Data Exchange Tests

In order to gain experiences with the PDF file format, tests have been carried out. The tests were done together with a group of twelve prepress and printing companies. The test started with creating a testform and are still going on. They will be finished when all problems including trapping and imposition with PDF are solved.

Testform

This testform is a A3 size (11.5 x 16.5 inches) QuarkXPress document consisting of 14 elements which come from different platforms and applications. The different elements were chosen to see, if PDF is able to solve difficult tasks, because the chosen elements sometimes lead in normal use to RIP problems. The 14 elements are the following:

1. A graphic file made with Illustrator 5.5 in which three images out of Photoshop are integrated: The file was stored as an EPS file and as such integrated into the QuarkXPress document. The file was done in CMYK and has a size of 45.1 MB. Some RIPs have problems to calculate this image because of its size.
2. A graphic file made with Freehand 4.0 consisting of Black and Pantone 1235: This file was used to test, if PDF is able to transport spot colors. It was stored as an EPS file and integrated into the QuarkXPress document. The file size is 31 KB.

3. An element containing two clipping paths where one is inside the other: Some RIPs have problems to cope with such a situation. The image and the clipping paths were done in Photoshop 4.0 and the background was added in QuarkXPress. The image is CMYK having a size of 7.6 MB.
4. A graphic file done on a PC in the application CorelDraw: It was then transferred to Macintosh by opening it in Illustrator 5.5 and stored as EPS file which was integrated into the QuarkXPress document. The colorspace is CMYK and the file size 307 KB.
5. A graphic file made with PowerPoint 3.0: It was stored from this application through the printer driver into an EPS file and placed into the Quark file. The color space is CMYK and the file size 407 KB.
6. A graphic file done in Illustrator 6.0: It was stored as EPS file and placed into the Quark document. The color space is CMYK and the file size 553 KB.
7. A graphic file done in Excel 4.0: Through the printer driver an EPS file was created which was imported into the Quark document. The color space is CMYK and the file size 345 KB.
8. An element consisting of type matter and fulfilling several functions: First seldomly used fonts were integrated in order to test that they are exchanged without problems. Secondly these fonts were trapped by 0.5 points in a blue background and in an image. It was tested, if these traps are retained when different data formats were used for data exchange. A blue background was established in the Quark document. The image was created in Photoshop 4.0. It is a grayscale image which was stored as a TIFF file. This element has a file size of 1.7 MB.
9 - 13. A series of images scanned with 300 dpi: The resolution of two images was downsampled in Photoshop 4.0 and different file formats were applied. The color space is always CMYK.

Element No.	File format	Resolution	File size
9	EPS	150 dpi	1.3 MB
10	EPS	300 dpi	5.1 MB
11	JPEG	300 dpi	422 KB
12	TIFF	150 dpi	1.0 MB
13	TIFF	300 dpi	4.0 MB

14. To control the output a control bar was integrated. It was created in CMYK and its file size is 30.1 KB.

Creating the PDF files for data exchange

Depending on what color data is generated in the QuarkXPress file, four different workflows are possible. The color data can be CMYK, RGB or CIELAB (see figure 2). If CMYK color data is chosen, two different PDF files are able to be created.

From the QuarkXpress document a PostScript file is generated. If in the QuarkXPress file CMYK data is used, the PostScript file can be preseparated or composite. If RGB or CIELAB data is used, only the composite case is possible. From these different PostScript files a PDF file is generated.

The preseparated CMYK PDF file consists of four individual monochrome files representing a single color of CMYK. If spot colors are used, they are generated in addition. The disadvantage of this workflow is that the document can not be controlled in color on the monitor. Only the single separations can be controlled in their monochrome representation. This way of exchanging color data is sometimes called «digital film» because the preseparated CMYK PDF files behaves like a piece of film. As the separation is already done and the screeening information is saved in the PDF file, the prepress company has full control over the output and is therefore able to overtake the full responsability for that file.

Unfortunately some RIPs overwrite the screening information and set the values of the in-house standard. In that situation the prepress company lacks of control over the output. If now the file is outputed with a different screen ruling, the dot gain settings may be wrong.

The advantage of this workflow is that the output is possible on all level of RIPs. Today still about half of the installed number of RIPs are PostScript Level 1 RIPs on which only this workflow is possible.

For the CMYK composite workflow a PostScript Level 2 RIP is needed, because the separations have to be made with the built-in RIP functionality of the RIP. To do this a special software is needed named Crackerjack from Lantanarips. This software sends the separation and the screening information down to the RIP where the PDF file is processed. The result are the four films or plates for CMYK. Crackerjack is able to separate spot colors, or to edit them into CMYK (see figure 3).

The third workflow is the color management workflow. It starts by introducing RGB or CIELAB colors into the QuarkXPress document from where the PDF file is created. Now is software needed to bring the ICC
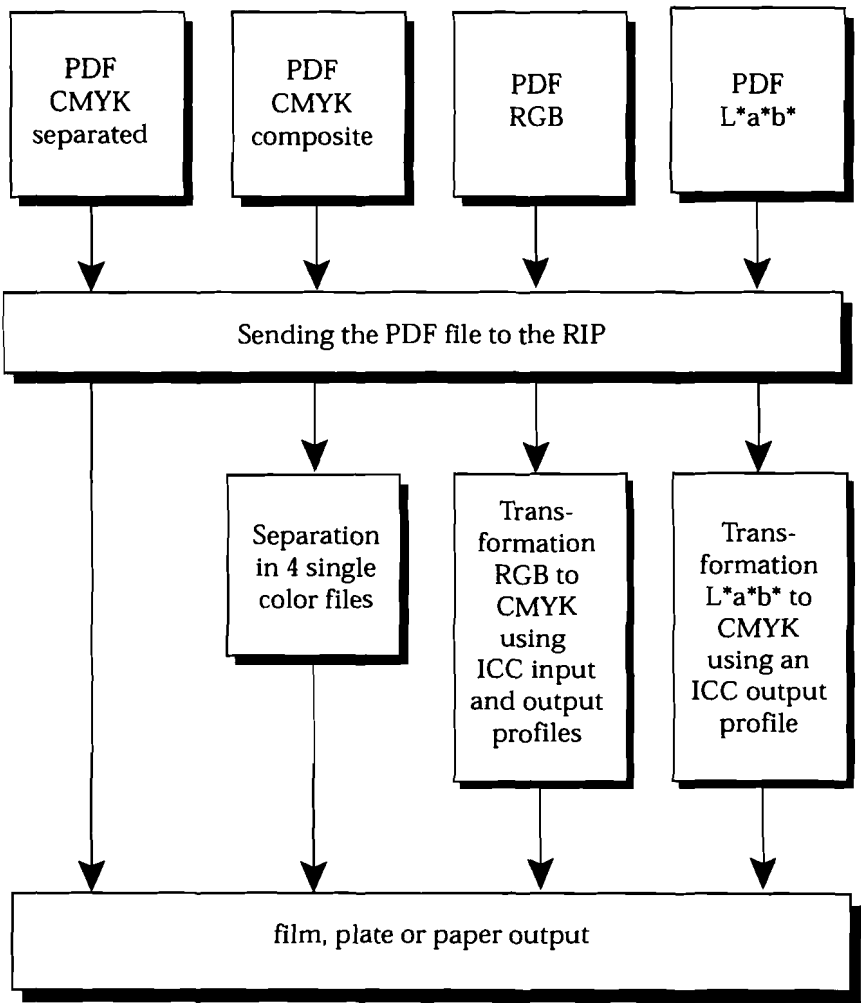
```
┌─────────────┐  ┌─────────────┐  ┌─────────────┐  ┌─────────────┐
│    PDF      │  │    PDF      │  │    PDF      │  │    PDF      │
│   CMYK      │  │   CMYK      │  │    RGB      │  │   L*a*b*    │
│  separated  │  │  composite  │  │             │  │             │
└─────────────┘  └─────────────┘  └─────────────┘  └─────────────┘
```

Sending the PDF file to the RIP

```
┌─────────────┐  ┌─────────────┐  ┌─────────────┐
│ Separation  │  │   Trans-    │  │   Trans-    │
│ in 4 single │  │ formation   │  │ formation   │
│ color files │  │  RGB to     │  │  L*a*b* to  │
│             │  │  CMYK       │  │   CMYK      │
│             │  │  using      │  │  using an   │
│             │  │ ICC input   │  │ ICC output  │
│             │  │ and output  │  │  profile    │
│             │  │  profiles   │  │             │
└─────────────┘  └─────────────┘  └─────────────┘
```

film, plate or paper output

Figure 2: Different PDF files and their output workflow

PDF CMYK composite

PDF CMYK separated

Cracker-Jack

Exchange

Exchange or Reader

Separation Conditions*

RIP Imagesetter

* - Screen ruling
  - Screen angle
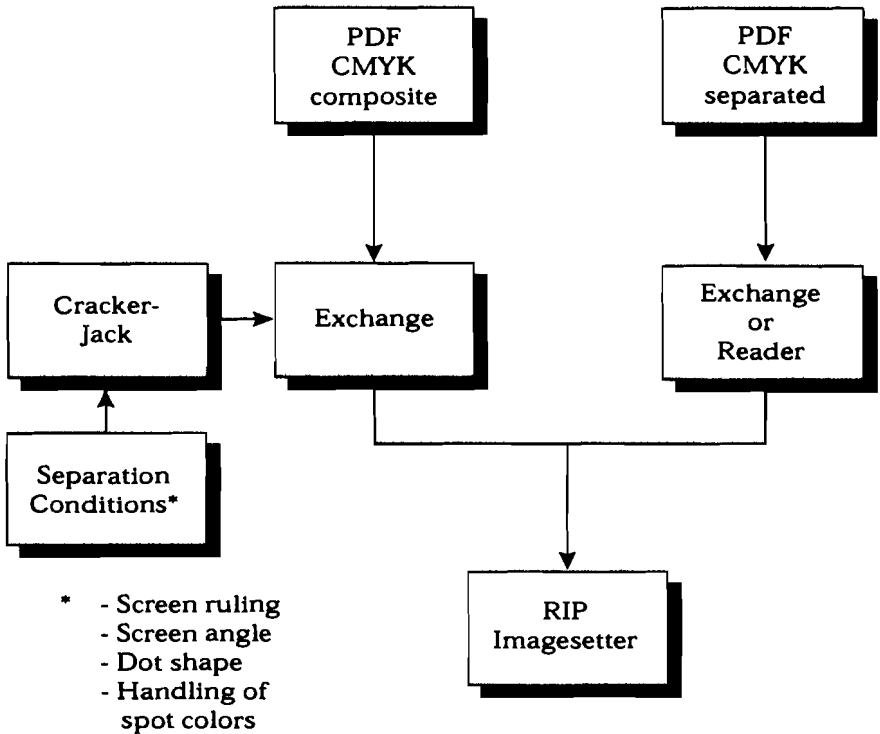 - Dot shape
  - Handling of
    spot colors

Figure 3: CMYK workflows with PDF files

input and output profiles into play. However, there is no such software on the market which works together with PDF files. Two software suppliers (Color Solutions, LOGO) have announced, that they will bring such software to the marketplace during the year of 1998.

EMPA has established instructions how these different PDF files are produced.

From the 14 elements of our testfile only one could not be converted into the PDF format file, i.e. the PowerPoint file. All other elements worked without any problems on all tested RIPs.

The proof should be made directly from the PDF file. In that way errors which are made while creating the PDF file can be dedected. From the native file only in-house proofs may be done, but the contract proof has to be made from the PDF file.

The job ticket is a part of the PDF standard, however, it has not been released until today. Therefore EMPA has created its one job ticket which contains all necessary infomation in addition to the digital data file. The job ticket itself is a PDF form which can be filled out in Acrobat Exchange and supplied to the printer together with the data file of the job.

The preflight checking may be done on both the native file and the PDF file. But more important is to check the PDF file prior to supply it to the printer.

Outputing the PDF file

If a printing company gets PDF files, a proof and a job ticket should be given with them. Once the proof and the job ticket are controlled, different workflows for the ouput are possible. The simplest case is when the document is directly printed from Acrobat Reader or Exchange. That is possible without any problems. But for certain jobs trapping and imposition is nessecary. There is software on the market with which it is possible to do this directly with the PDF file.

Today a PostScript Level 2 or a PostScript 3 RIP is needed to work with the composite workflow. But there are software suppliers working on new tools allowing to do the same output on PostScript Level 1 RIPs.

With PageMaker 6.5 it is possible to work in the color management mode. This requires to embed the input profiles into the images. PageMaker 6.5 is able to identify these profiles and they are transferred to the PDF file. The output profile is transformed into a CRD (color rendering dictionary) which is downloaded to the RIP. In a this way RGB and CIELAB PDF files can be converted to a CMYK output.

CRDs are no longer state of the art. A ICC output profile should be applied which, however, requires that new software has to come to the market. Today such software is available to attach profiles to PS and EPS files. The software manufacturers have announced that similar solutions will be possible for handling PDF files (see figure 4).
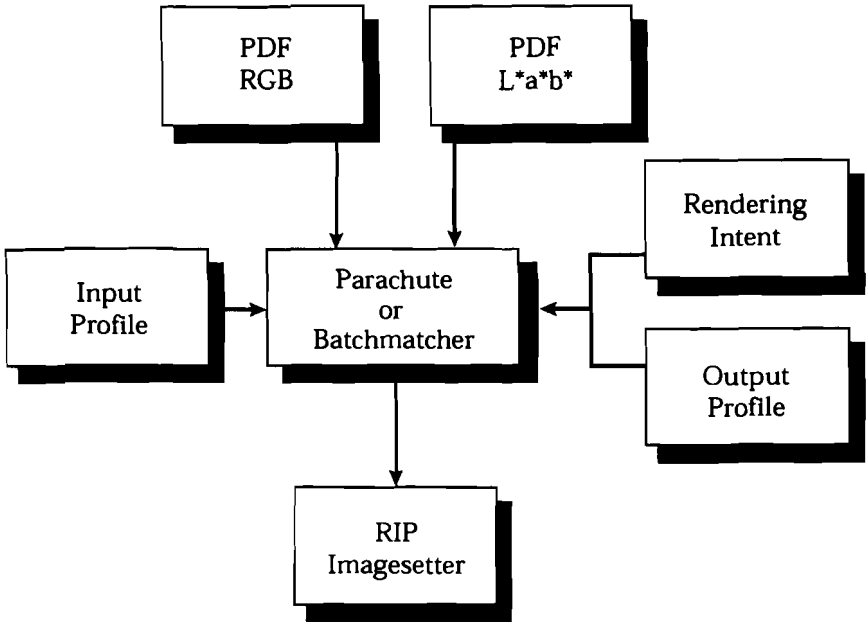
Figure 4: Color management workflow with PDF files

This will allow to attach the input profiles, the rendering intent, the output and the simulation profiles to the RGB or CIELAB PDF file. The color space transform will be calculated either on the workstation or in the RIP.

PDF files can easely be brought to the Internet. One has to decide whether the resolution, which was originally chosen for printing, should be downsampled. If the file is planned to be downloaded and be used for printing, the high resolution must be kept. Otherwise the resolution has to be downsampled to 72 dpi.

The future of PDF

CGATS and ISO are doing specifications for the use of PDF files for data exchange. Therefore PDF will be the future data exchange format. Moreover, it is used widely in the Internet. Maybe PDF will be a universally used file format in future.

Soon there will be software on the market to create PDF files directly out of QuarkXPress. PageMaker is able to do this already with its version 6.5. Therefore no PS file has to be created in order to obtain a PDF file.

There are still some minor failures in both the underlying PDF specification 1.2 and the important Adobe Acrobat software. But there will be updates of the specification and the software products. It is forseeable that the problems which exists today, will be solved very soon.

A bigger problem to overcome is to bring the composite workflow to the PostScript Level 1 RIPs. There are software companies working on that subject. Such software will have a big market, because most of the installed RIPs are still PostScript Level 1 RIPs.

The consequent application of the present or any other specification standardising the use of PDF will help to make the data exchange more secure and help to save a great amount of time and costs.