DATA FORMATS IN DIGITAL PREPRESS TECHNOLOGY

Lidija Mandic [1], Sonja Grgic [2], Ida Srdic [3]

[1] Faculty of Graphic Art, Getaldiceva 2, 10000 Zagreb, CROATIA, e-mail: mandic@grf.hr
[2] Faculty of El.Eng. and Comp., Unska 3, 10000 Zagreb, CROATIA, e-mail: sonja.grgic@fer.hr
[3] Croatian Telecom, Jurisiceva 9, 10000 Zagreb, CROATIA, e-mail: ida.srdic@ht.hr

Abstract: *The printing industry is on its way to become a multimedia industry. Printing will not be the only media and means of outputting informations. Printed products and online media have a common prepress stage (premedia) which only diverges just before the functional output as electronic media or as printed media. A good example of the potential of the multimedia market is the use of efficient workflow. The data file can be directly transmitted to different media by efficient workflow. Thanks to digital technology, all captured text and image data can be used for multimedia purposes. Incorrect data format can effect the quality of a prepress product. The aim of this article is to discuss the most prevalent data formats used in open prepress systems.*

Key words: *Images, Data Formats, Digital Prepress Technology*

1. INTRODUCTION

Digital prepress technology is confronted with a multitude of data formats. A few device components of a digital publishing system and the most prevalent data formats used in open prepress systems are shown in Fig. 1. Data from text, graphics and picture-capturing devices form the stream of the input data, [1]. The output side is dominated by the page description language PostScript. Being device independent, the data structure of the Postscript can be used to control the complete range of output device.
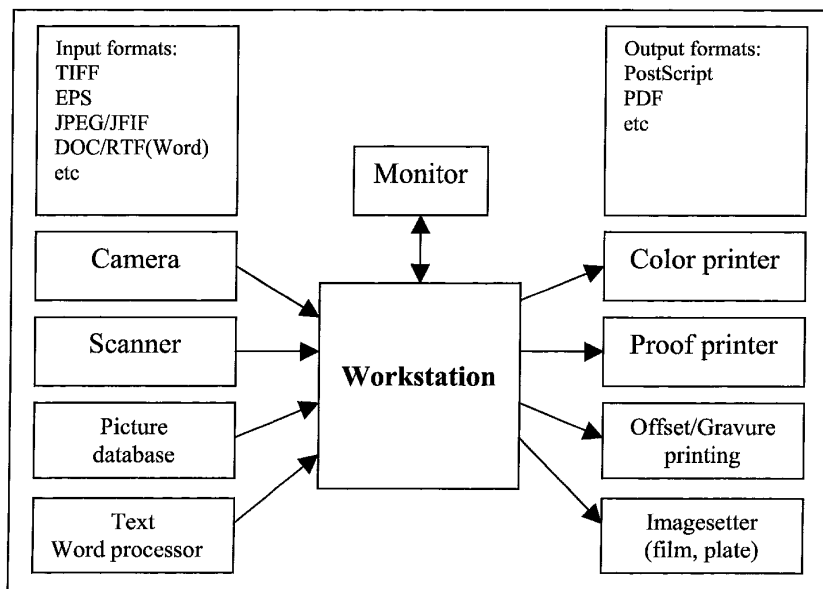

Fig. 1. Digital publishing system

The core of the possible workflow is a data base containing all the data required for production. Layout, text, graphics and pictures are still the main components of printed information and must be made available in edited form. The basic stages in the creation of a digital page are shown in Fig. 2.
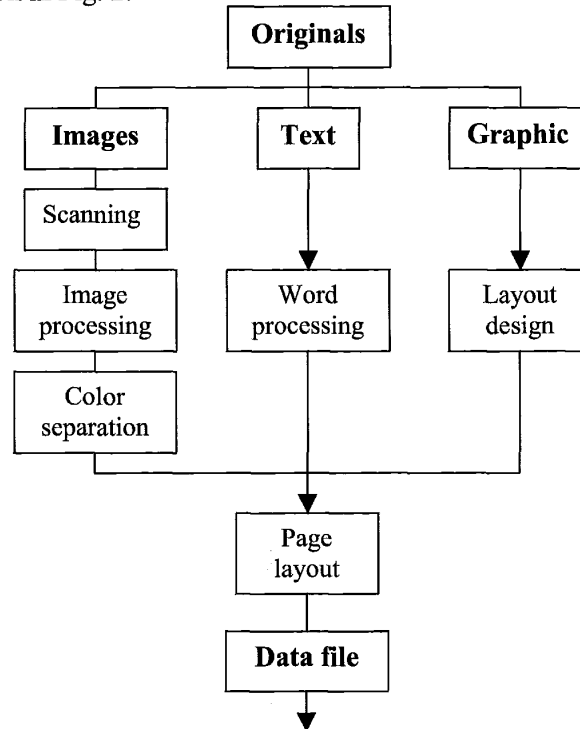
```
                          ┌──────────────┐
                          │  Originals   │
                          └──────┬───────┘
          ┌──────────────────────┼──────────────────────┐
   ┌──────┴───────┐      ┌──────┴───────┐      ┌──────┴───────┐
   │   Images     │      │    Text      │      │   Graphic    │
   └──────┬───────┘      └──────┬───────┘      └──────┬───────┘
   ┌──────┴───────┐             │                     │
   │  Scanning    │             │                     │
   └──────┬───────┘             ▼                     ▼
   ┌──────┴───────┐      ┌──────────────┐      ┌──────────────┐
   │    Image     │      │    Word      │      │   Layout     │
   │  processing  │      │  processing  │      │   design     │
   └──────┬───────┘      └──────────────┘      └──────────────┘
   ┌──────┴───────┐
   │    Color     │
   │  separation  │
   └──────────────┘
                          ┌──────────────┐
                          │    Page      │
                          │   layout     │
                          └──────┬───────┘
                          ┌──────┴───────┐
                          │  Data file   │
                          └──────────────┘
```

Fig. 2. Digital page from individual elements

The text data are prepared in "Word" format and edited in a layout program. Images are usually scanned in, or digitised for publication. Image-editing programs generate bitmaps, also called raster images. A digital image is a rectangular array of dots, or picture elements, arranged in m rows and n columns. The expression $m{\times}n$ is called the resolution of the image. When an image is digitized, its appearance depends on the resolution - that is, the number of pixels per linear unit - as well as on the number of bits that are used to represent each pixel. Each pixel is assigned a specific location and color value, [2]. The number of bits that are used to represent each pixel also helps to determine image quality. Bitmap images is resolution-dependent, it contains a fixed number of pixels to represent its image data. As a result, a bitmap image can lose details. All picture data are generally saved in TIFF format or, to save data transfer times, compressed in JPEG format. Graphics constitute the third main element of a printed page. These data are usually saved in the form of vector-based data files, which cannot be edited or positioned in a layout program. Vector graphics is resolution-independent, so without loosing detail or clarity. Graphics are saved in EPS format. Layout programs are software packages allowing for flexible work and for integrating the elements (text, images and graphics) on pages. The page components text, images, graphics and layout, must be prepared and organised for further stages in processing, in order to avoid errors or breakdowns in the workflow. System must be able to access the original resources, so a layout file is supplied that contains the definitions of special color or information regarding color separation.

2. DATA FORMAT

2.1 Input data

Bitmap formats are used to exchange and save picture data. They are mostly saved in TIFF, EPS and JPEG/JFIF formats.

TIFF is a classical, pixel based data format, originally developed by Aldus (now Adobe Systems) and Microsoft, [3]. It is used to exchange files between applications and computer platforms. The basic parameters of a picture (e.g. resolution) are saved in standard "tags". TIFF is a strong, extendable data format, based on the "Lempel-Ziv-Welch" compression method (LZW), making it a simple, no-loss data compression option. It supports alpha channel. Alpha channels store selections as 8 bit grayscale images and store masks, which allow manipulating, isolating and protecting specific parts of an image.

EPS (Encapsulated PostScript) data format allow the saving of vectors. It is used to transfer PostScript language between applications, [4]. Graphics data files are mainly stored as outline-based data. An EPS file can contain both outline and bitmap-based data. EPS is a special file version of the PostScript page description language. It is supported by page-layout programs. EPS data format does not support alpha channel.

The JPEG (Joint Photographic Expert Group) File Interchange Format is used particularly for large data sets, optimised prepress workflows. The JPEG picture data compression is lossy compression technique that provides the best results with continuous-tone images, [5]. The JPEG format is specified both for RGB and CMYK data, but not for CIELAB data. The loss rate of this method depends on the content of the picture.

2.2 Output data

The data generated with the different layout programs are not PostScript files, but they contain all the information required, in line with PostScript conventions.

The page description language PostScript presently still plays the dominant role in the output of a digital publishing system. PostScript is a general, interpreter-based computer programming language, which was developed by Adobe founders John Warnock and Charles Geschke. It is used to describe the appearance of a document page and to transfer it to an output system. Postscript can code all elements of a document, including text, graphics and pictures. It is generally generated in the driver of the computer system, transferred to the printing device via a network, and translated by the interpreter (RIP) into the language of the output device. Device independence is ensured by the internal data structure (vectors and Bézier curves), [6]. PostScript includes both the device-independent and device-specific components. One variant of the PostScript file is the "Desktop Color Separation" format (DCS). If a single color separation of a previously color-separated PostScript (existing in CMYK color mode) is sent to an image setter, this would normally imply that the entire file has been sent to the image setter, including all the other color separations. In DCS files, the four color separations are saved separately in four separate files, a fifth file containing a low-resolution, non-separated version of the image is used for layout purposes. In the so-called "separated workflows", the color separations and their trapping are calculated completely, saved separately, and transmitted via the network as separate files.

The PDF (Portable Document Format) is a data format used to describe documents. PDF document stores each page of a publication separately, so the file need not to be completely interpreted in order to print or display the contents of a page. Since PDF files have already been RIP'd, they can ensure a far more reliable result when printing or imaging.

Table I shows some characteristics of input and output data formats.

Table I Comparison between data formats

INPUT DATA FORMAT		
	TIFF	EPS
structure	linear	programming language
file format	binary	ASCII
reliability	yes	no
compression	LZW	JPEG
masking	alfa channels (PS)	curves (clipping path)
graphics	possibility of saving curves	yes
modifying in layout program	yes	no
DCS	no	yes
defining of screening	no	yes
OUTPUT DATA FORMAT		
	PostScript	PDF
structure	programming language	data format
modifying	yes	no

3. CONCLUSION

Processing the pre-separated data would not make sense in the new, media-independent production environments and workflows. Data formats have both the advantage and the disadvantage. Pixel-based data format (TIFF) preserves picture data, they can been modified in layout program, but required a large storage space. The outline-based data format (EPS) can be resized, the considerable data compression rates can be achieved, but the contents of the data can no longer be modified. On the other hand, unwanted changes can be prevented. While PostScript constitues a complete programming language, PDF is a data format. Contrary to PostScript file, a PDF document stores each page of a publication separately. The separated objects can be represented on the screen, and not simply in their ASCII code, as in a PostScript file.

REFERENCES

[1] E. Giorgiani, T. Madden, *Digital Color Management*, Addison-Wesley, Harlow, 1998
[2] I.A. Witten, A. Moffat, T.C. Bell, *Managing Gigabytes*, Academic Press, San Francisco, 1999
[3] Aldus, *TIFF Revision 6.0*, 1992
[4] B. Walroad, *QuarkXPressUnleashed*, Random House, 1994
[5] D. Salamon, *Data Compression: The Complete Reference*, Springer, New York, 2000
[6] Adobe Systems Inc, *PostScript Reference Manual*, Addison-Wesley, Harlow, 1992