TIFF/IT

Tag Image File Format for Image Technology (TIFF/IT) is an ANSI standard (and an ISO draft standard) that builds on the Aldus 6.0 version of TIFF and carries forward the work done on DDES® (Digital Data Exchange Standards®) and IFEN (Intercompany File Exchange Network).[1] TIFF/IT provides a media independent transport mechanism for raster images. The resulting standard is an interesting meshing of high-end and desktop publishing formats.

[1] For more information on DDES, DDAP, IFEN, ANSI, and ISO, please refer to these Linotype-Hell Technical Information articles: *Digital Advertising* (pages 49-54 of the 1994 notebook) and *Graphic Arts Standards* (pages 33-36 of the 1993 notebook).

In practice, TIFF/IT should make it easier to trade data between high-end systems from different vendors, but it should also simplify data exchange between high-end systems and desktop environments (and vice versa). It should also play an important role in the success of direct-to-plate and direct-to-press technologies.

What is TIFF?

TIFF is a format for storing and interchanging raster (as opposed to vector) images. Typically this refers to data that comes from scanners, frame grabbers, as well as paint and photo-retouching programs. TIFF is capable of describing images in a number of formats and also supports several compression methods. It is not tied to any proprietary products and is intended to be portable to a variety of computer operating systems, file systems, compilers, and processors. And, it is designed so that it can evolve gracefully as new functions become necessary.

It is the tags[2] from TIFF which made the format particularly attractive to the supporters of DDES. These tags provide a more flexible file structure that allowed them to keep much of the early DDES work, even though DDES was written specifically for image transport on 9-track magnetic tape. In essence, the file headers used on magnetic tape became tags which could be used for a variety of transport media. Because TIFF had been designed with evolution in mind, it was possible to create new tags for TIFF/IT to satisfy requirements of high-end systems.

[2] Remember, TIFF stands for *tag* image file format.

TIFF

The TIFF concept began late in 1985 when Aldus Corporation met with a variety of desktop scanner vendors to discuss format standards. Their consensus was that none of the existing proprietary formats had the industry clout to become the de facto standard. In addition, page layout programs were likely to be burdened with the task of reading a wide variety of proprietary formats. A standard scanned image format specification would greatly simplify the process of including scanned images into page layout documents created by programs like Aldus PageMaker.®

Early drafts were circulated and by August of 1986 the first version of TIFF was approved. Microsoft played a role in the drafting of TIFF, and later formally endorsed TIFF for Windows.™ This helped establish TIFF in both Macintosh® and Windows applications, and also opened the door for entry into other platforms.

TIFF/IT background

While Aldus was finalizing the TIFF 6.0 specification in the spring of 1992, the IT8 committee of the American National Standards Institute (ANSI) was working on TIFF/IT (officially known as ANSI IT8.8). TIFF/IT was approved in November of 1993.

TIFF/IT incorporated these three elements of DDES:

- **IT8.1** – User Exchange Format (UEF00) for the Exchange of Color Picture Data Between Electronic Prepress Systems via Magnetic Tape (DDES00)
- **IT8.2** – User Exchange Format (UEF01) for the Exchange of Line Art Data Between Electronic Prepress Systems via Magnetic Tape (DDES00)
- **IT8.5** – User Exchange Format (UEF03) for the Exchange of Monochrome Image Data Between Electronic Prepress Systems via Magnetic Tape (DDES00)

While these standards were tied to distribution via magnetic tape, the intent of TIFF/IT was to use these data formats but make them media independent. In the paragraphs below you will see how IT8.1 became CT, IT8.2 became LW, and IT8.5 became MP, BL, and BP. Two additional component, HC and FP (see below), were borrowed from IFEN.

IFEN

The IFEN data formats were specified by Crosfield, Linotype-Hell, and Scitex. Later, these companies commonly proposed these formats to the ANSI IT8 committee for integration into TIFF/IT as a subset. This is an important point because it illustrates that the high-end system vendors are willing to propose and support standards for the general benefit of the graphic arts community. IFEN was introduced in 1990 at DRUPA.

TIFF/IT components

The TIFF/IT format is made up of three primary components. Only the first component (CT) is actually part of the Aldus TIFF 6.0 specification:

- **Contone image (CT)** – Each pixel is described by four bytes, one for each of the four process colors (eight bits per pixel per color): cyan, magenta, yellow, and black (CMYK). This format is equivalent to TIFF CMYK which was previously known as TIFF Class S (separated TIFF).
- **Linework image (LW)** – High resolution, contone, multi-colored graphic and text elements described as run length compressed data. LW is superimposed onto CT during color separation. While LW pixels may be assigned an opaque color, they may also be transparent (to let the background CT show through).
- **High resolution contone image (HC)** – Run length encoded format which is commonly used to describe high resolution edges between contone images. Like LW pixels, HC pixels may also be assigned a transparent value. The resolution of this format must be high to avoid stairstepping at the edges of masked images.

As part of the appendix to the specification, another component is described:

- **Final page (FP)** – Describes a complete page formed by superimposing CT, LW and HC information. FP comes directly from work done in IFEN.

LW, HC, and FP have no equivalent formats in Aldus TIFF. Why is this? These file formats are unique to raster-based file structures of high-end systems. PostScript™ linework is usually reproduced with vectors, not rasters. In addition, PostScript has no use for the HC format because masks are generally described using vector clipping paths. Given that LW and HC formats are not used in PostScript, it would be superfluous to include them, or the FP format, in TIFF since TIFF was originally conceived to exist in a PostScript world. But, it does makes sense to include them in TIFF/IT because high-end systems make use of them.

There are also three monochrome or binary components of TIFF/IT. These come from work done in DDES (IT8) and "are similar to their color counterparts except that their formats take advantage of the reduced amount of data associated with monochrome (single color) and binary images."[3]

[3] Source: ANSI IT8.8-1993, page 7.

- **Binary line art (BL)** – Binary line art (or run length encoded bitmap) image or file. Each pixel is represented by a single 'value'.
- **Binary picture (BP)** – Binary picture (or byte packed bitmap) image or file. Each pixel is represented by a single bit.
- **Monochrome picture (MP)** – Monochrome picture (or continuous tone) image or file. Each pixel is represented by a single byte.

TIFF/IT, P1, & ISO

The ISO draft standard (CD 12639, #1, TIFF/IT) proposes two levels of conformance: TIFF/IT and TIFF/IT P1.

- **TIFF/IT** "is intended to support a transport independent means for the exchange of various images used in the prepress, printing and graphic arts fields, typically between CEPS"[4] (color electronic prepress systems).
- **TIFF/IT P1** "is also intended to support a transport independent means for the exchange of various images but for a broader range of fields to include prepress, printing and graphic arts and additionally DTP and information processing fields…TIFF/IT P1 is a simpler and more straightforward means of exchange."[5]

[4,5] Source: Committee Draft of International Standard - CD 12639 (#1) Graphic Technology - Prepress digital data exchange - Tag image file format for image technology (TIFF/IT), page 5.

TIFF/IT allows the exchange of files with a wide variety of characteristics. P1 removes some of the flexibility of TIFF, but makes those files simple to read. In essence, P1 has more restrictions which make it simpler for receiving, but more complex for sending. P1 uses the most common desktop defaults within TIFF. While a TIFF/IT P1 reader would be simpler to create, it might not be able to read all TIFF/IT files.

Compression methods

Before going any further in the discussion of TIFF/IT, it is important to look more closely at some compression methods. Some data formats are very closely linked to the methods used to compress them. In fact, with the high-end DDES formats, if some method of compression had not been available, their large file sizes would have made them impossible to use.

A variety of compression methods are used in TIFF and TIFF/IT:

- **Run length** (sometimes abbreviated as RLE, run length encoded) – Run length data compression methods decrease file size by encoding sequences of identical symbols. For example, in a bilevel image, the image is composed of only two types, black and white. In some cases, there may be sequences that include long strings of either black or white characters. Instead of encoding a sequence of 20 white pixels as: WWWWWWWWWWWWWWWWWWWWW, this string could be encoded as 20W. Two examples of run length compression methods are: Pack Bits and CCITT Group 3 1-dimensional Huffman coding. PackBits is a byte-oriented run length compression scheme commonly used with Apple Macintoshes. CCITT Group 3 is a coding method whereby frequently occurring items are given shorter codes than less frequently occurring items. Bilevel images may be compressed with either Pack Bits or CCITT Group 3. Grayscale can't be compressed using CCITT Group 3. Run length methods are lossless, which means that there is no loss of quality due to compression. (This is in contrast to lossy methods where data may be lost, see JPEG below.)
- **LZW** – LZW stands for Lempel, Ziv, and Welch. (These are the names of the compression scheme's creators.) LZW is designed to be able to compress all kinds of data, including images of a variety of bit depths.[6]

[6] LZW works particularly well on screen dumps.

LZW is lossless and works quite well on bilevel images, however, for grayscale and full color images (particularly those images with a lot of detail) LZW may not be able to offer significant amounts of compression.

LZW is an adaptive compression method which means that the compression technique is dynamically adjusted based on the content of the data being compressed. LZW may also be termed a dictionary method of compression because it creates dictionaries which are used to compress commonly repeated data.

LZW has been patented by Unisys and must be licensed to be used, even though many developers mistakenly believed it to be in the public domain.

- **JPEG** – The Joint Photographic Experts Group (JPEG) was formed to create a standard for color and grayscale image compression. (JPEG is effective only on continuous-tone color spaces.) JPEG describes not one, but a variety of algorithms, each of which is targeted for a particular class of applications. These algorithms fall into two classes: lossy and lossless.

  Lossy – These algorithms are based on transformation into the frequency domain using the discrete cosine transform (DCT)[6], plus Huffman coding of the frequency components. Though they involve some loss of reconstruction accuracy, they provide substantial compression without significant image degradation.

  Lossless – These algorithms use an algorithm based on a two-dimensional differential pulse code modulation (DPCM) technique combined with Huffman coding of the differences.

The LW and HC formats are two particularly good examples of data formats that use lossless run length coding as a compression method.

LW and HC

While high-end systems generally use raster formats alone, in the PostScript world both raster and vector formats are common. And, LW files are not line art in the PostScript sense of the word. In PostScript documents, line art is created as PostScript code (which describes the lines, curves, and fills of a piece of line art.) This vector PostScript line art is resolution independent (i.e., it can be output on a variety of PostScript devices and still produce the best quality that the device can achieve).[7]

Historically, traditional high-end systems work with raster data only. This means pixels, and lots of them.[8] Scanned images contain many pixels, each of which can represent a range of colors. These scanned continuous tone images are relatively large in size, but because they are ultimately halftoned, the scan resolution does not have to be extremely high.[9] Line work (since it includes items with well-defined edges like text, rules, and logos) needs higher resolution to avoid stairstepping. To merge the line work convincingly with the contone data, not only does the spatial (i.e., scan) resolution have to be high, but the tonal resolution (i.e., bit depth) also has to match the continuous tone scan. This results in a big line art file. As an example, a 300 pixel per inch, 1 square inch, CMYK file requires about 350 KB of storage space. By increasing the resolution to 1800 pixels per inch (about what a line work file requires), the file size increases by a factor of 36 to 12.4 MB. (An 8 x 10 inch file at that resolution would be nearly a gigabyte in size!)

Luckily, LW files, because they contain repetitive areas of the same color, are easily compressed using run length encoding methods. With compression, the LW size for a page is typically reduced to 1 to 5 MB. The same is true of the HC files used for high resolution edge definition. They also are of great size, and they benefit greatly from compression. The size of LW and CT files is one obscure technical reason why PostScript has had such great success. For most desktop applications it just made a lot more sense to create line work with vectors than to create enormous raster files.

[6] DCT is an algorithm that converts the data (pixels) into sets of frequencies. The first frequencies in the set are the most meaningful; the latter, the least. During compression, the less meaningful frequencies are quantized with less pixels or even stripped away based on allowable resolution loss. Source: *The Computer Glossary, 6th ed.*

[7] Sometimes in the PostScript world, one bit scans are also referred to as line art, but again, this differs from the high-end definition of line work.

[8] The ANSI TIFF/IT specification refers to the components of an image as a "bag of pixels".

[9] For more information on scan resolution, please refer to these Linotype-Hell Technical Information articles: *Scanned File Size* (1992 notebook) and *Scan Resolution & Sharpening* (pages 43-46 of the 1994 notebook).

However, in defense of the LW format, it should be noted that LW provides a simple method to describe as a single object a multicolor element with complex shapes. In PostScript such an element would have to be composed of several objects, one for each color, and each one to be described by its own vector outline. For very complex shapes (which are typical in high quality advertising and catalog work) the vector approach may result in huge PostScript files and long RIP times. (For example, imagine the complexity of a PostScript vector file describing a shape like the silhouette of a tree with all its branches, leaves, and the open areas between leaves).

The LW format can easily be viewed on a monitor without the need to run it through a RIP. In addition, some last minute changes (i.e., color changes, shape editing, and cut & paste operations) are simpler to do in the LW format. And, they can be done independent from the application which generated the LW. Once those corrections are made the file is immediately ready for direct film exposure without a RIP. These are some of the reasons why TIFF/IT formats are meaningful as an alternative page description method, mainly for complex high-end work.

Conclusion

So what we see in TIFF/IT is a collection of formats, some with roots in the high-end, some with roots in the PostScript world, all joined together through a common file structure. Why did the industry think that it was necessary to bring these elements together in a standard? The main reason is ease of interchange. But, TIFF/IT also allows powerful high-end formats that take advantage of DDES to break the bonds of magnetic tape. TIFF/IT could also simplify direct-to-plate and direct-to-press operations by providing a stable raster format which is desirable for a number of reasons:

- In a direct-to-plate and direct-to-press environment it is much more costly to have a PostScript error occur late in the process. A stable raster file can be passed to a recorder with less chance for error.
- Imposition, trapping, color correction, and adjustments for press conditions can often be handled more easily with a raster file.
- Particularly for complex high-end work, it may be advantageous to have the linework represented as a raster file rather than as a vector file. This is a very important issue for ghosting and image collage where images are supposed to show through other images.
- It is generally easier to estimate the output time of a raster file based on its size. When PostScript vectors are RIPed, even though the file size may be small, the job still may take a long time to output.

Acknowledgments

Many thanks to Mary Abbott of NPES, Steve Carlsen, Dr. Jim King, and Steve Zilles of Adobe Systems, Alan Darling of Quad/Graphics, Mark Jones of Ringier America, and David McDowell of Eastman Kodak for their help in producing this document. I would also like to thank the following people from Linotype-Hell: Wolfgang Boppel, Ray Cassino, Ted Cooper, Vern Kellie, Dr. Dieter Preuss, and Dennis Ryan.

Appendix A: TIFF image types

The latest version of TIFF is revision 6.0 which was completed in June of 1992 There are four baseline image types described in the TIFF 6.0 specification: bilevel, grayscale, palette color and full color:

- **Bilevel** – Contains only two colors: black and white. (This image type was called TIFF Class B in earlier versions.)
- **Grayscale** – Can contain multiple shades of gray. (This image type was called TIFF Class G in earlier versions.)

- **Palette color** – Has only one component per pixel which is used as an index into a lookup table called a color map. The lookup table contains numerous RGB values. For an 8 bit RGB image, the lookup table contains 768 values.[10] These 768 values are used to represent 256 RGB colors. (This image type was called TIFF Class P in earlier versions.)

[10] $3 \times 2^8 = 768$ (3 represents the three colors, RGB, 2^8 equals 256)

- **Full color (RGB)** – Each pixel contains three components (red, green, and blue). There is no lookup table. (This image type was called TIFF Class R in earlier versions.)

In addition to the baseline image types, other color spaces supported include: CMYK, colorimetric RGB, YCC, and CIE L*a*b*:

- **CMYK** – This is a color separated format in which each pixel consists of four components (C, M, Y, & K). There are, however, ways to include extra colors or non-color information (like opacity). TIFF CMYK is also called TIFF Class S.
- **Colorimetric RGB** – Most RGB is device specific, but, by adding a method for describing and characterizing the RGB data it is possible to provide calibrated RGB values that relate to the CIE 1931 XYZ color-matching functions. (CIE stands for Commission International de l'Eclairage.[11])
- **YCC** – TIFF YCC uses the YCC color difference color space that is popular in digital video applications and is also used by Kodak in Photo CD. TIFF YCC is also called TIFF Class Y.
- **CIE L*a*b*** – The CIE L*a*b* color space is colorimetric and perceptually uniform. It is particularly useful for manipulating continuous tone images. CIEL*a*b* is used in LinoColor.™

[11] For more information on CIE, please refer to these Linotype-Hell Technical Information articles: *Color Spaces and PostScript Level 2* (1992 notebook) and *Color Gamut* (pages 41-44 of the 1993 notebook).

And finally, there is also a TIFF for facsimile transmission. (This is also referred to as TIFF Class F.)

Appendix B: Ordering TIFF 6.0

TIFF 6.0 may be ordered in hard copy for $25 (plus $4 shipping and handling) by calling 800-831-6395. It may also be downloaded from Aldus forums on CompuServe, AppleLink and the Internet:

- **CompuServe** – Enter Library 10 of the ALDUSSP forum. (Macintosh users: Download TIFF6D.SIT. PC users: Download TIFF6.ZIP.)
- **AppleLink** – Path–Third Parties: Aldus Developer Folder and TIFF Files: Final TIFF 6.0 Specification.PS (PKG).
- **Internet** – Available via public ftp on sgi.com in the directory graphics/tiff (file name tiff6.ps).

Appendix C: Ordering TIFF/IT

The ANSI TIFF/IT standard (ANSI IT8.8) is available at a nominal charge through NPES The Association for Suppliers of Printing and Publishing Technologies (703-264-7200). An order form is also available through the NPES fax on demand system (800-874-0858).

Please direct any questions or comments to: Jim Hamilton, Marketing Department, Linotype-Hell Company, 425 Oser Avenue, Hauppauge, NY 11788
(For subscription information on the Linotype-Hell technical information series, please call 1-800-842-9721.)

October 1994, Part Number 7041

© 1994 Linotype-Hell Company. All rights reserved.
• Linotype and Hell are registered trademarks and LinoColor is a trademark of Linotype-Hell AG and/or its subsidiaries.
• Aldus and PageMaker are registered trademarks of Aldus Corporation.
• Digital Data Exchange Standards and DDES are registered trademarks of DTI.
• Macintosh is a registered trademark of Apple Computer, Inc.
• Microsoft is a registered trademark and Windows is a trademark of Microsoft Corporation.
• PostScript is a trademark of Adobe Systems Inc. which may be registered in certain jurisdictions.
• The Digital Distribution of Advertising for Publications and DDAP are trademarks of the DDAP Association.
All other company and product names are trademarks or registered trademarks of their respective owners.