

TIFF, the Image Storage Standard

Kenneth R. Rohr

Abstract—In products like imaging systems, where archived information must be accessible for many decades, adhering to standards is absolutely essential. Many generations of new computer systems will be used during the life of the archive. They will all need to read your images. Keeping images in TIFF files enables such a migration strategy.

I. INTRODUCTION

IMAGING systems are very inexpensive when compared to the value of the information stored. A \$20-40K scan station can easily archive over a million pages per year. The cost to replace *just one year's* images will exceed the cost to replace the entire scan station.

How does a customer protect his investment? How can a service bureau efficiently support many different customers? Customers and service bureaus both need standards for interchange of images.

Imaging systems designed for multi-vendor computer environments use TIFF files. This image format was defined in 1986 by a group of seven major vendors of scanners and related software, including Aldus, Hewlett-Packard, MicroSoft and Microtek. In addition to the original authors, TIFF is now supported by hundreds of computer companies and independent software vendors.

II. TIFF FILES

What is TIFF? The acronym means the Tagged Image File Format. It is

- Extensible—new image types can be introduced without invalidating older types.
- Portable—it is independent of hardware and operating system types.

The insides of a TIFF file consist of two main parts: a header and the image data. The header

consists of a series of “tags”. Each tag is a small piece of data that identifies things like the resolution of the image, the width and length of the image, and so on. Some tags, like the name of the page, may be ignored gracefully by programs that don't care about that data. In this way new tags can be added to the TIFF spec and most old programs will continue to work as well as they ever did. They just ignore any new tags that they don't know about. New programs might offer some enhanced feature by reading and using the new tag.

One tag identifies the format of the image data itself. The image data may be uncompressed, or it may be compressed using one of the five different compression schemes currently supported by TIFF. As technology progresses, more compression types will likely be added. Another tag identifies which “byte-order” (this accommodates different CPU hardware). Not all programs are required to support all compression schemes and both byte-orders. To establish compatibility between two imaging systems, they need to have at least one compression scheme and byte-order in common. The choices are:

Compression	Byte order
Group 3 - fax Group 3-mode 2	Motorola
Group 4 PackBits LZW (for color)	Intel

This may seem confusing, but it's as simple as the way we write dates. November 3, 1993 can be compressed using several “compression schemes”, e.g. 11/3/93 or Nov. 3, '93. To be unambiguous, we also need to know the order of the information. 11/3/93 in England means March 11, whereas in the U.S. it means November 3. As long as we identify which

conventions we are using, there's no problem communicating the information. That's exactly what TIFF does with the tags.

III. THE CURRENT SITUATION

With all this support, portability and extendibility, one might hope that all imaging systems use TIFF files. But some, especially older architecture, imaging systems use proprietary structures to hold images. Images are buried in database fields (called BLOBs), or bundled into large files with some proprietary format that only the original vendor knows about. To retrieve images, you must use a retrieval station from that vendor. This is a lock-in strategy that constrains the customer from buying any imaging systems from other vendors. It also makes a service bureau buy a scan station from that vendor in order to support his customer.

Before you buy any imaging system, you should ask whether the system stores images in TIFF files. A service bureau will want to be in the business of converting paper to TIFF (also microfilm to TIFF). If you're buying a system

for a service bureau that cannot easily create TIFF files at full scanner speed, you will be limited in the customers that you can support efficiently. If you're the system's end user, the situation could be even worse. You may be forced to buying products from just a single vendor because your images may not be viewable on any other equipment. These are both unacceptable predicaments in a field that will change over the life of an image archive.

Acordex provides efficient systems that use TIFF as the native image storage with scanning capacities of 1,500 to 3,300 pages per hour, continuous operation. Investing in such an imaging system can meet your current productivity goals and position you well for the long term future.

The author is with Acordex Imaging Systems (978 352-5500), a product development and system integration company specializing in high performance document management systems. <http://www.acordex.com/~cordant>