

Digital Image File Formats and their Storage -- TIFF, JPEG & JPEG2000

February, 2010, v20 © Tim Vitale [18zq>zzn]

1 - Introduction	1
2 - Image File Formats	2
TIFF vs JPEG vs JPEG2000	2
TIFF File Format	2
TIFF with the LZW and ZIP Lossless Compression Options	3
BigTIFF Format	3
JPEG File Format	3
JPEG Compression: Examples	4
Figure 1: JPEG Compression Examples	4
Figure 2: TASI Graphic on JPEG Compression	5
JPEG2000 File Format	6
Figure 3: JPEG2000 compression Examples	6
Additional JPEG2000 Information	7
3 - Storage of Digital Image Files	7
Figure 4: Failure Trends in Large Disk Arrays	7
Storage Recommendation	8
IT Department	8
Internal and External Hard Drives	9
Longevity of Hard Drives	10
RAID Array Systems	10
CD and DVD Optical Storage	10
Hard Drive Storage	11
Additional Digital Storage Information	11

1 - Introduction

The storage of image information is crucial for its long-term preservation. While digital images can be stored indefinitely without deterioration, they can be lost through neglect. A digital file can be permanently "lost" if stored without regard for basic computer technology (not backed-up) or on inappropriate media (optical disks).

The recommended storage medium is the harddrive (HDD), which are viable for 3-10 years, with an average of about 5 years. Although any HDD can fail at any time, it is usually backed-up on another HDD or in an internally-redundant RAID array (mode 1 or 6, not 5). Optical media (CD±R, DVD±R) fail without warning in 2.5 to 25 yrs; their readers (DVD drive) probably won't be available in 15-20 yrs.

TIFF is the preferred archival image file format. The TIFF format is a file wrapper that holds image data directly from the imaging processing software, with all the color data for each pixel, including all the tags necessary to reconstruct the image as made in any good image processing software [Photoshop].

The next preference-level down uses file compression. There are two types of compression: lossless and lossy. There are lossless compression file formats that save some space (20%+) while not changing the image data. Examples of lossless are TIFF-ZIP, TIFF-LZW and JPEG2000 (in lossless color-saver mode). Lossy compression is epitomized by the JPEG format, which does harm to the image information by design.

Lossy compression of an image file diminishes the potential of the spatial and color image information by throwing information away to save space and improve download speed. Compression should be used if the original image data is not as important as

- **the space it occupies**
- **the speed of download**
- **the speed of movement over a local area network (LAN)**

Compression should not be a default operation, but a choice that is considered fresh for each project.

Lossy compression is most effective at reducing file size and increasing download speed. Both JPEG and JPEG2000 (in lossy mode, not color-saver mode) encode the original RGB image data permanently altering the original numerical data. Figures 1, 2 and 3 show the degree of damage lossy compression makes on spatial and color information.

The JPEG2000 lossy compression engine (using wavelet technology) produces images with superior appearance to the older JPEG (DCT) compression technology. However, no matter the vast superiority of JPEG2000, support by web browsers for the format is still limited. Only Safari on Mac has full support for JPEG2000; the PC version of Safari needs a Quick-Time plug-in. Wikipedia has

tables that show support for JPEG2000 by imaging software http://en.wikipedia.org/wiki/JPEG_2000, and browsers at <http://www.fnordware.com/j2k/jp2samples.html>.

2 - Image File Format

The file format is critical to the preservation of an image. The preservation standard is the TIFF file (tagged image file format) because it just wraps structure around the digital information that is used within the image processing software. It is a file wrapper, not a specific format that makes the information conform to its protocol. TIFF holds all the (a) spatial (resolution/detail), (b) color (RGB or Lab color numbers for each pixel), (c) ICC/ICM profiles, (d) metadata and (e) preservation information required to create a digital master of the original image. When opened in good quality software the image can be recreated on your screen the way it was on screen when saved. The JPEG2000 compression engine can be used to reduce image file size by roughly 30% to 75%, when used in the color-saver mode [lossless mode (LL). The JPEG format was intended for use when the space a file uses has more value than the information it holds; use of the format by digital camera manufacturers was an unintended use of the tool; it was not designed not file storage, as was JPEG2000.

TIFF vs JPEG vs JPEG2000

The TIFF format holds all imaging data in a sound and archival manner. Proper permanent storage of digital objects requires multiple copies of a file, and preferable, one held in a different geographical location. One TIFF file is large, three are much larger. Organizations (LoC and NARA) with large numbers of TIFF files have been looking for a format that saves space while preserving image fidelity.

The JPEG format is a “lossy” compression protocol that decreases file size from 100 to 4000 times by permanently removing image information. The compression changes spatial resolution, tonal range and color. It is not a preservation format and was never meant to be one; it is a very good tool for facilitating image use on the web as it was designed to be.

JPEG2000 is the next generation of image compression tools, using “wavelet” technology. It is a superior tool. It is often called lossless, but this term only applies to one of the operational modes, the color-saver (LL) mode. At the present level of support by web browsers (Mac Safari) it will not replace the original JPEG format any time soon. JPEG2000 offers improved image quality over JPEG, but the file size can be larger. The **fnord** software website (makers of a JP2 plug-in for Photoshop) has an informative direct side-by-side comparison of a 16KB JPEG and a 16KB JPEG2000 compression, made from the same image <http://www.fnordware.com/j2k/jp2samples.html>, showing better outcome for the JPEG2000 version. Even after 10 years of implementation, JPEG2000 is still in the “initial” phase; widespread adoption is still very slow. The reasons are unclear, but lack of commonly available encoding software and browser support seem a probable reasons.

TIFF File Format

A TIFF file (Tagged Image File Format) is a **file wrapper** that contains all the sections, elements and tags required to hold bitmapped raster or vector images in any word order, as a grayscale (B&W), RGB, CMYK, CIE Lab and YCbCr (color space) file; other properties include:

- **Uncompress 1-, 8-, 12-, 14-, 16- & 32-bit images**
- **Compression possible: lossless ZIP & LZW, and, lossy JPEG & JBIG**
- **Any Resolution: (1-10,000+ ppi)**
- **Metadata of all types**
- **Support Mac and PC digital word order**
- **ICC profiles stored for use**
- **32-bit architecture**
- **4 GB file size limit**
- **255+ tags (v6): 25 Baseline tags, including pixel density, image copyright, date and time; 65 Extended tags, white point, JPEG tags and YCbCr coordinates; 80+ Private tags, including Photoshop options and ICC profiles; 55 EXIF tags, camera technical metadata; 30 GPS tags, geographical location metadata**

The virtue of a file wrapper is that it holds original data in the original digital word order and format. The value of the TIFF format is that it can contain a perfect version of the capture, with full resolution and color, pixel-by-pixel. Data is held uncompressed with relevant metadata about the technical creation parameters, including device and settings, image visual description terms, copyright data, and file preservation parameters. The TIFF format does not require a license. It is dependent on the software that processes the image.

The specific software used to make the TIFF image file is critical because it's a wrapper rather than a standard or protocol. Some brands of imaging software make TIFF files may not be able to be

opened by other software because of [their] proprietary methods of handling digital language or machine code such as tiling, row layout, cell dimensions, etc may not be support by another software package. Adobe Photoshop is considered the best image processing software package; Apple Aperture is also considered high quality software.

Faulty image capture and damage to the image done during processing by the operator will be reflected in the saved TIFF file. If a file made in "XYZimagetool" software will not open in Photoshop, it is because the off-brand software compromised the image data. The file will need to be opened using that software, and then exported using a file format that is common to both software, which does a s little harm of the data as possible.

TIFF-header and TIFF-tag readers are available to check the integrity of an image file that won't open or has other problems. TIFF file tags can be viewed on the Windows platform using **TiffTagViewer**, downloaded from <http://www.awaresystems.be/imaging/tiff/astifftagviewer.html>. A simple search will yield others.

A PowerPoint presentation on [Getting the Most from a TIFF Image](#) (2005) can be found at http://cool.conservation-us.org/coolaic/sg/emg/library/2005-06-vitale-documentation-tiff-image/2005-06-vitale-documentation-tiff_files/v3_document.htm. It shows how to make and adjust an image to get the most relevant color information possible, and where the inherent faults in color management are found.

TIFF with the LZW and ZIP Lossless Compression Options

TIFF-LZW compression is the lossless type. However, it will only decrease the file size by a small amount (6% in tests) or even increase the file size by 50% in one test using a non-photo image.

TIFF-ZIP is also lossless compression protocol. In tests, it did a better job of compressing photographic image(s) than LZW. The rate of file compression was about 20%, down to zero-compression for a non-photographic test image.

The [Still Image Compression Research](#) webpage <http://videopreservation.conservation-us.org/tjv/index.html> has a series of compression experiments using these protocols, near the bottom of page, on both photographic and non-photographic images.

BigTIFF – Next Generation

The current limitation of a TIFF image file is that it uses 32-bit offset architecture, which means it is limited to a maximum of 4 GB file size. Up until about 2004, Photoshop (v7), could not handle greater than 2 GB of RAM, and could not open or create images with more than 90,000 pixels in one direction. This limitation was broken in 2005 (CS1 or later) so the TIFF file wrapper can hold image files large than 2 GB.

The next generation of the tagged image file format is BigTIFF <http://www.awaresystems.be/imaging/tiff/bigtiff.html> with 64-bit architecture. BigTIFF closely resembles the TIFF format. The existing TIFF libraries can easily extend their support to the new BigTIFF variant.

All the properties of TIFF are still present in BigTIFF. All known tags are being used in BigTIFF. All supported bit-depths and data types remain valid. The arbitrary number of extra channels, tiling and striping schemes, variety of optional compression schemes and private tag schemes that made TIFF useful in pre-press, storing scientific imaging data, preservation and other applications, all remain intact.

JPEG File Format

Images saved in lossy compression formats such as JPEG are not meant for high-resolution imaging, nor, was the format created to be used for image preservation. The JPEG file format was created by the Joint Photographers Expert Group (started in 1986) and released in 1992. The stated intention was to allow for the quick access and display of images on the web; the web was still quite new in 1992.

Use as a primary imaging format by compact camera and dSLR cameras **was not** an intended goal of the joint photographers expert [group] committee. The JPEG format was adopted by the camera manufacturers to compensate for the small size of the early memory cards (8-32 MB) that were used to store images onboard a camera [as film]. In an era of 4 to 32 GB+ memory cards, the JPEG format is not necessary for storage Camera Raw image files that only range in size from 6 to 12 MB.

JPEG is a lossy compression process. For a given quality setting (1-12), different images will yield widely differing file sizes and image appearances. An image with a lot of texture and fine detail will produce a large JPEG file, while one consisting only of blue sky will be very small.

JPEG compression uses the Discrete Cosign Transfer (DCT) coding algorithm. The compression process is built on the following steps:

- 8-bit RGB (only) support, forces higher bit rates to 8-bit [possible 2:1 compression]
- Conversion to YUV/YCbCr color space, 128 steps rather than 256 per channel [compression]
- YUV (4:4:4) values are downsampled to 4:1:1 [3:2 compression]
- Spatial compression in 8x8 blocks of 8-bit pixels [approx. 2:1 compression]
- High frequency data zeroed out of the 8x8 DCT blocks [image data lost]
- 8x8-DCT block coefficients quantitized to variable levels of compression [variable compression]
- Entropy coding to reduce repetitive Hex codes, called Huffman coding, based on a zigzag pattern using variable length codes and integers [additional level of variable data compression]

Lossy compression is an irreversible way of reducing the size of data by approximating it from the original bitmap image. Once image information has been lost it cannot be recovered, except by starting the process "fresh" from the original bitmap. Trying to improve the appearance of a JPEG image by re-compressing at a higher quality setting achieves very little except an increase in file-size.

JPEG Compression: Examples

A 68 MB TIFF color file, that is compressed to a 40-60 KB JPEG file will be compressed about 1350:1 through the process. Much of the detailed spatial image information was lost in the process, but the JPEG file is still good enough for viewing on a monitor. Figure 1 shows the damage done to both color and spatial information, the lighter the color the more damage that has been done.

CRT monitors (cathode ray tube) are imperfect analog output devices with resolution equivalent to about 72-96 dpi and a tonal range of 35:1, or about $5\frac{2}{3}$ f-stops of light. Not as much tonal range as a color transparency (3.6 Dmax), about equal to a traditional photographic print, but less than a good inkjet print that can have a Dmax of 2.2-2.4 ($7\frac{1}{3}$ f-stops).

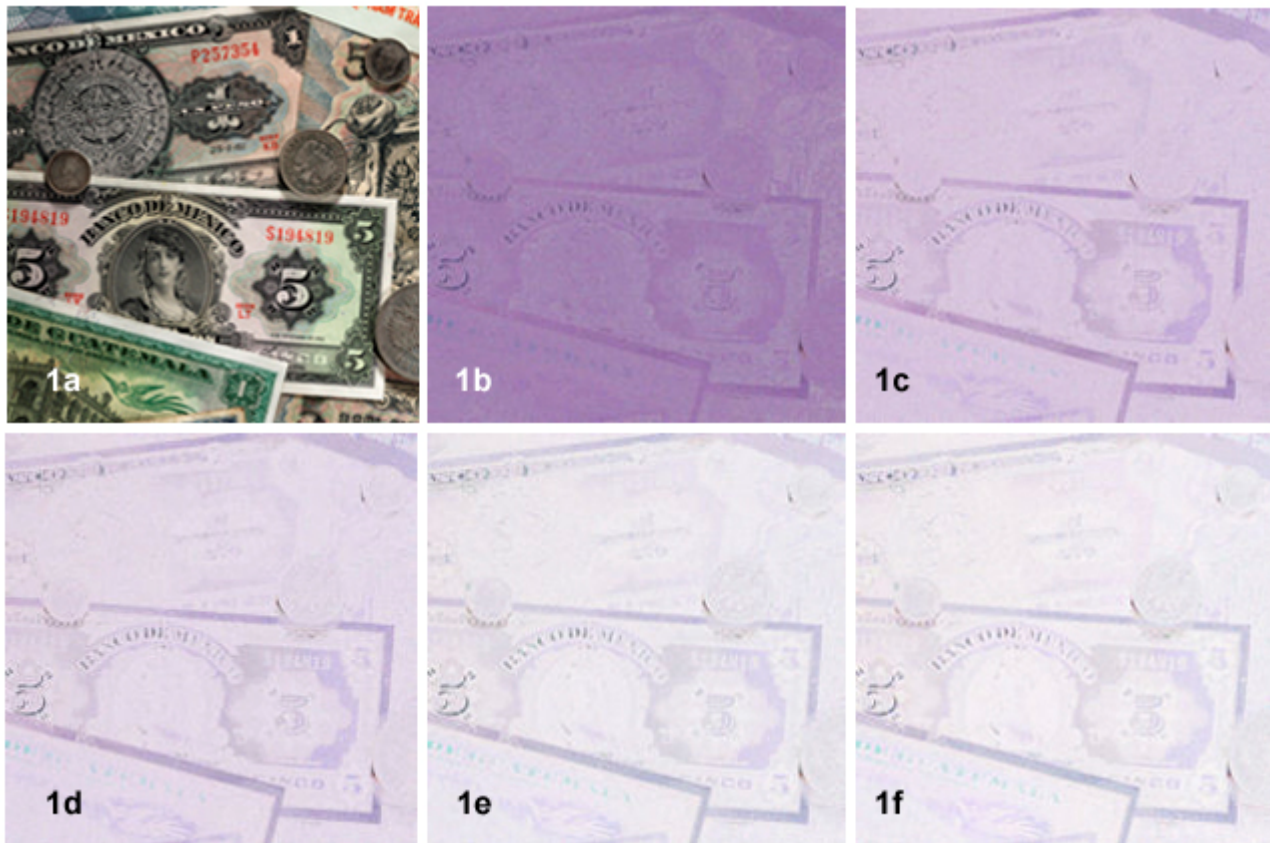


Figure 1: JPEG Compression Examples. The images depict the level of damage done to the original image by the various JPEG compression quality levels. **Fig 1a** is the original; **Fig 1b** shows the difference between the original and the compressed image at level 12, the highest quality; **Fig 1c** shows level 11; **Fig 1d** shows level 10; **Fig 1e** shows level 8; **Fig 1f** shows level 6. Notice that the more compression the lighter the difference-image. More light means a greater difference created during the [Image] <Apply Image...> part of the image difference [subtraction] process.

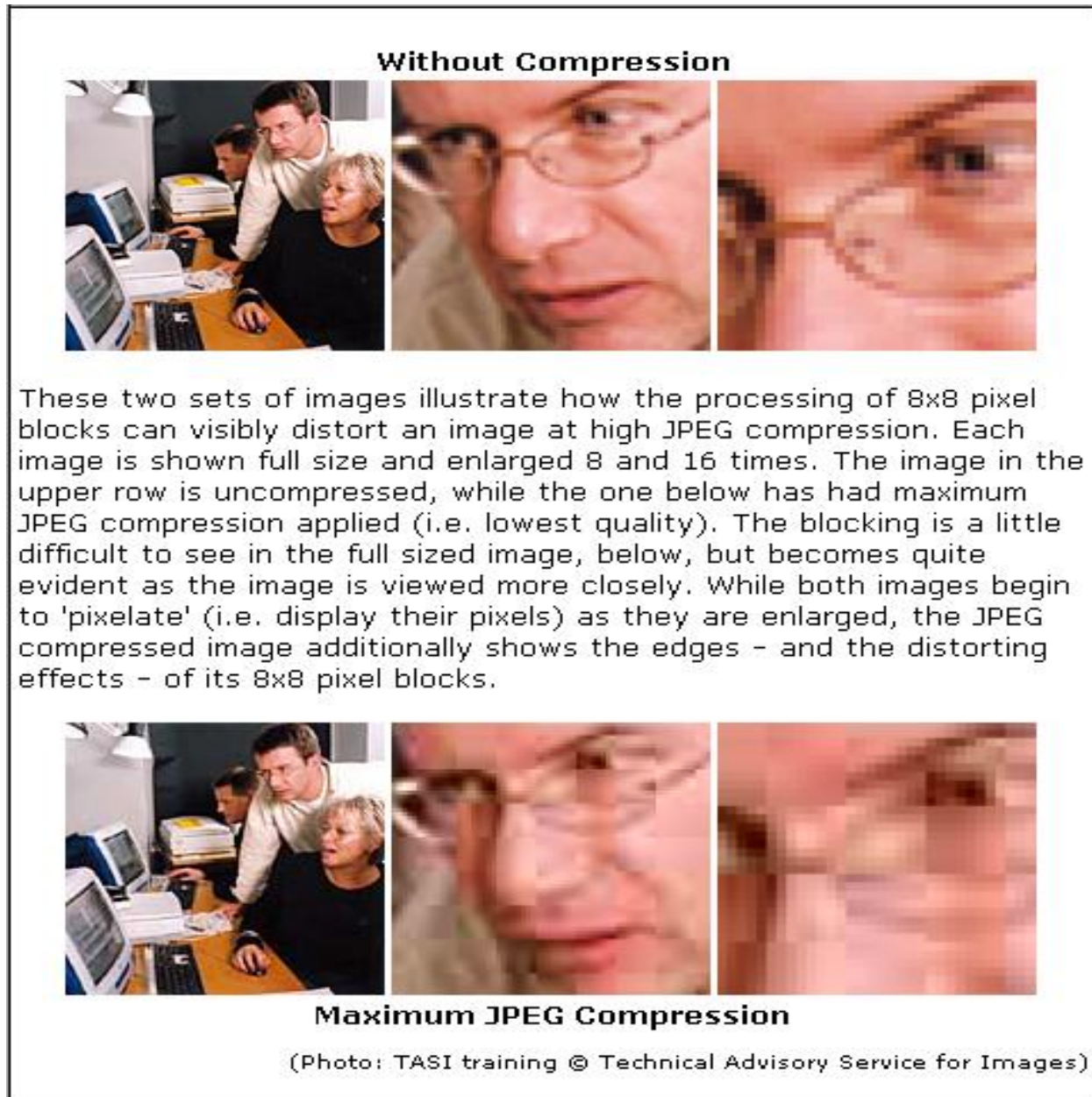


Figure 2: TASI Graphic on JPEG Compression. The graphic above was pulled from the TASI website. It shows the degree of spatial resolution changes before and after maximum JPEG lossy compression. Note the small blocks on the lower right and large blocks in the lower center.

As a means of comparison, the average B&W or color negative film has the potential for about $5\frac{1}{3}$ f-stops of light, or a density range of 0.2 to 1.7 D, while most transparency film have a tonal range of 3.6 D (≈ 3.9 Dmax). An original image before compression has the tonal range of the device, which can range from 5 f-stops (1.65 D) for most 8-bit imaging systems, up to, $11\frac{2}{3}$ stops (3.9 D) in 14/16-bit imaging systems. JPEG images have an 8-bit limitation which means $5\frac{1}{3}$ stops of light.

Many of the better LCD monitors have higher resolution (above 96 ppi) depending on settings. Modern LCD displays can have a tonal range of 300:1 ($7\frac{1}{3}$ f-stops of light or 2.4 D), up to 10,000:1 (4.0 Dmax or $12\frac{1}{3}$ f-stops of light) for the most expensive (\$4k) LCD displays.

Prints from a (high-resolution) inkjet printer have a resolution of about 1200-1440 dpi for matte papers and 2400-2880 dpi for glossy papers. At 1440 dpi, the resolution is 10 to 15 times greater than the resolution of a standard monitor. The typical tonal range of an Epson Ultrachrome inkjet print is about 2.1 to 2.4 Dmax, or 7 f-stops of light.

JPEG2000 File Format

An image saved using the JPEG2000 protocol is destined for compression. However, depending on settings it can be either lossless (LL) or lossy (Lo) compression. JPEG2000 lossless compression goes through an encoding process, thus, in any mode, the format is storage protocol not a file wrapper. After compression, the compression ratio and file size ranges (roughly) as follows

- 1.5:1 to 3:1 -- for Lossless compression (about 30% to 70% compression)
- 100:1 to 1000:1 -- for Lossy compression, at various levels of quality

An image is encoded and saved. When the image file is viewed, it is decoded by viewing software. Most web browsers (with the exception of Safari in MAC) are not JPEG2000 capable. The biggest problem with the format is support by web browsers.

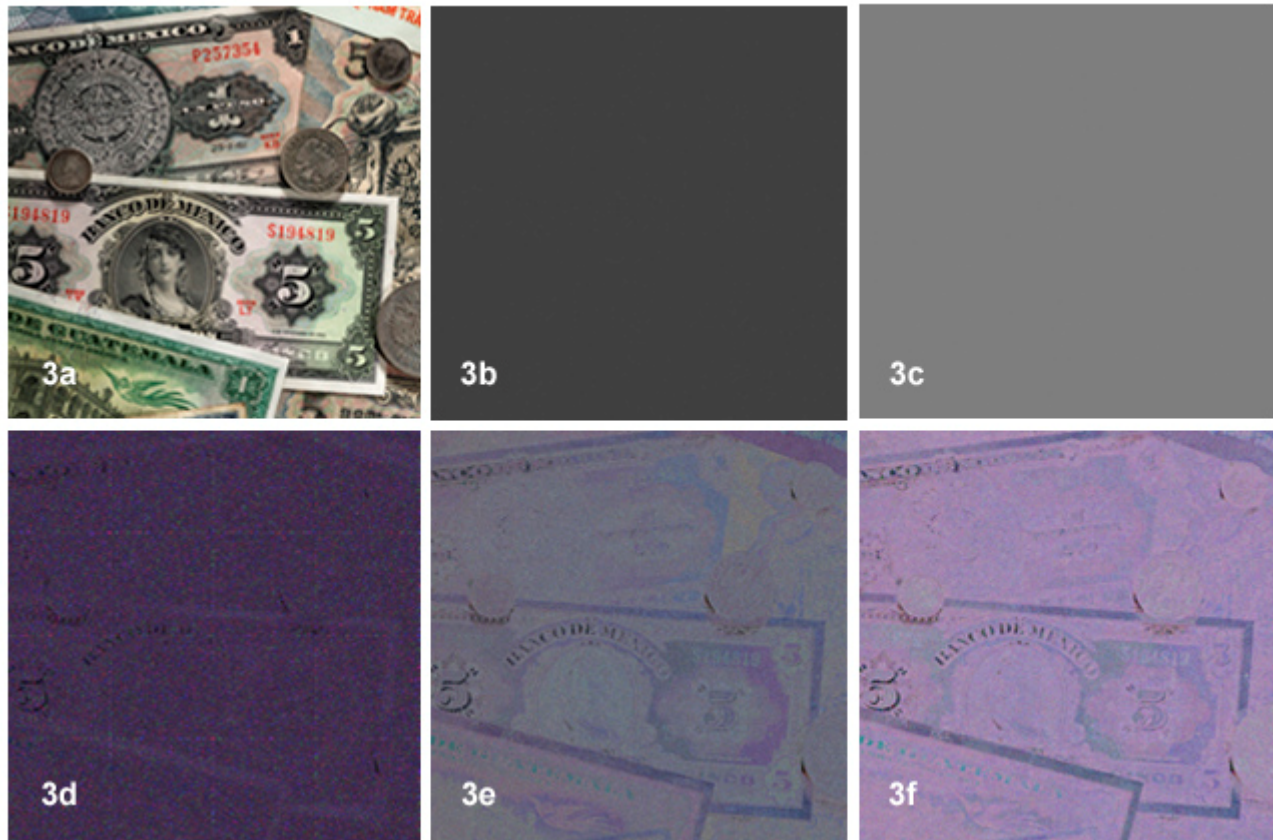


Figure 3: JPEG2000 compression Examples. The images reveal the presence and level of damage done to the original image by lossless (LL) and lossy (LO) the various JPEG2000 compression levels. **Fig 3a** is the original; **Fig 3b** shows the difference between the original and the compressed image using lossless (LL) compression, the continuous black tone shows no changes for a 37% compression of file size; **Fig 3c** shows the same "difference" image with the black level adjusted to gray to reveal any small changes, none are present; **Fig 3d** shows lossy compression at quality level 75 (Lo75), the purple color reveals an overall shift that is quite similar, the slight grain in the tone show unevenness in the color shift; **Fig 3e** shows lossy compression at the quality level 50 (Lo50), obvious changes based on color; **Fig 3f** shows lossy compression at quality level 25 (Lo25), changes in color are obvious. A detailed explanation with image before and after files that can be downloaded can be found at <http://videopreservation.conservation-us.org/tjv/index.html>.

The lossy mode of the technology converts each image into a spectrum of image tiles with specific spatial and color properties. The original spatial and RGB numerical value information are converted into data that can be compressed. In the lossless mode, the re-converted image is indistinguishable from the original; see Fig 3b & 3c. In lossy mode the image damage appears to be less than JPEG compression. In yet another mode, the original image data can be saved and then delivered in "portions," at full resolution, upon request; the overall file size is large than the TIFF equivalent. The encoding process follows these steps:

- Data Ordering
- Arithmetic Coding
- Coefficient Bit Modeling
- Quantization
- Wavelet Transformation

• DC Level Component Transformation

The most recent versions of Photoshop (CS 3 & 4) now support JPEG2000 without plug-ins, although there are other brands of J2K plug-ins, such as **fnord** <http://www.fnordware.com/j2k/> and LEADTOOLS <http://www.leadtools.com/sdk/compression/jpeg2000.htm> that can be downloaded for used in most versions of Photoshop.

Wikipedia has two tables (at different URLs) that show software support for JPEG2000 by image processors http://en.wikipedia.org/wiki/JPEG_2000, under the heading **Application Support**, and by web browsers at <http://www.fnordware.com/j2k/jp2samples.html> under the heading **Image Format Support**.

Additional JPEG2000 Information

The official JPEG200 website can be found at <http://www.jpeg.org/jpeg2000/>. The IEEE review paper by Marcellin, et al, **An Overview of JPEG-2000** (1999) can be found at http://www.rii.ricoh.com/~gormish/pdf/dcc2000_jpeg2000_note.pdf. Basic information can be found at http://rii.ricoh.com/~gormish/pdf/dcc2000_jpeg2000_note.pdf, which has 24-page overview of the format.

An early version (free) of the **ITU specifications** on JPEG2000 can be found at <http://www.jpeg.org/public/fcd15444-1.pdf>. More modern versions of the ITU specifications must be purchased through standards organizations.

There are fervent followers of the technology http://www.oreillynet.com/pub/a/javascript/2003/11/14/digphoto_ckbk.html, while others have called the value of JPEG2000 into question <http://graphicssoft.about.com/gi/dynamic/offsite.htm?zi=1/XJ&sdn=graphicssoft&zu=http%3A%2F%2Fwww.levien.com%2Fgimp%2Fjpeg2000%2Fcomparison.html>.

Many workers agree that Kakadu is the best software for making JPEG2000 image files. A demonstration version can be downloaded from http://www.kakadusoftware.com/index.php?option=com_content&task=view&id=26&Itemid=22. The former versions of the KDU software were a command line only software package; it now has a GUI for all the common operating systems. While this is an advance over command line operation, the software is quite difficult to use. The behavior is pedantic and doesn't seem to remember previous actions, so each file save requires many navigational steps. In addition, the usage is not "Windows like" meaning there is a learning curve, fortunately the scope of the operation is small so a half-an-hour will suffice.

3 - Storage of Digital Image Files

The safest and most cost effective method of storing image files is to use a harddrive. A modern HDD will have a useful life of about 5 years, by then its size will become so impossibly small that the data will be migrated to a new HDD to save space. Based on average MTBF data (mean time between failure), file migration will occur long before a HDD will fail. Even though MTBF data suggests otherwise, backup of primary storage is always recommended because the drive could fail before the predicted MTBF. Discussion of early HDD failure can be found, but today's HDD are significantly different from those made just 5-10 years ago.

In a 2007 PC World article by Scheier, he reviewed research into HDD failure at server farms http://www.pcworld.com/article/129558-2/study_hard_drive_failure_rates_much_higher_than_makers_estimate.html. It was found that heat did not correlate with failure rate as it has been suggested. Based on Google's review of SMART (Self-Monitoring Analysis and Reporting Technology) data from their numerous HDDs, there is no one predictor of failure. In [Pinheiro, Wolf-Dietrich and Barroso](#) (2007, at the USENIX Conference), it was reported that failure rates were higher than predicted by manufacturers, from 1.7% (1 yr) to 8.6 % (3 yr). However, [Scheier](#) (above) reports an annualized failure rate (AFR) of only 0.88%, less than one percent.

In the past, CD±R were considered suitable for archival storage. Life expectancy (LE) predictions for optical media vary widely and are based upon multiple factors including: data density, layer bonding, reflective & dye layers and writing speed. DVD±R's have few independent performance records, but they are certainly not as stable as CD±R because the density of data is about 6-times higher and they are not using the most stable dye layer.

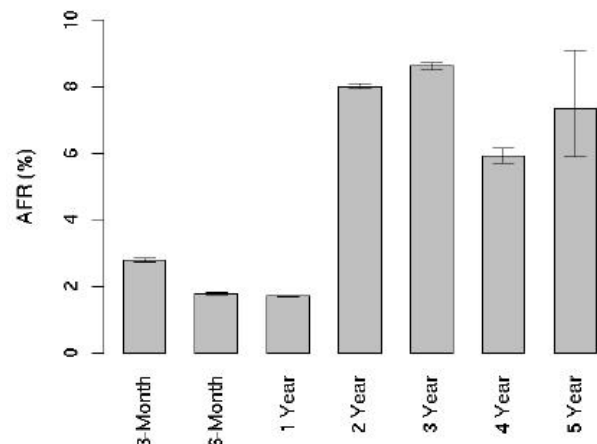


Figure 4: Pulled from [Failure Trends in a Large Disk Drive Population](#) by Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz André Barroso (Google Inc.); see Figure 2 on p4 of the 2007 PDF.

The bottom line is that optical disks hold only relatively small amounts of data for the effort required to make and certify the disks for archival use. A 1-TB HDD now sells for under \$100 (2 TB OEM Seagate can be found for \$100 online), thus HDD are much more attractive storage options because even a dual layer DVD±R (8.4 GB) is a small fraction of a 1-TB HDD, and DVD-R_DL have been shown to fail earlier than most DVD [yet to be published a new *Restaurator* article by Joe Iraci]. A spindle of 25 dual layer (DL) DVD-R will cost about \$30 for 210 GB of storage, this is 1.5-times the cost of similar real estate on a modern 1 TB HDD. In time CD±R and DVD±R drives will become obsolete, just as floppy drives are now rare on new computers.

Optical media is not as easy to migrate as an external HDD. For an external HDD, one just plugs in or docks the HDD, after it is recognized the drive is opened automatically. The target folder(s) are just dragged across the screen to the intended new home for the data. The user can just walk away while the file transfer proceeds, often taking several hours for hundreds of GB of data. No additional action is required upon completion; file verification is integral to the transfer process.

Storage Recommendation

Purchase a one new external HDD every year (for each computer or device needing backup); it will be twice as large as the year before and cost less. Backup last year's HDD on this year's drive. For internal HDDs, migrate data from older drives to a larger drive(s) every 2 to 3 years. Modern NAS (network attached storage) drives that sit on the local LAN can be configured to backup new data every night, much like large IT enterprises that continually backup their system every night, although those procedures generally call for a complete backup of every file. A 2 TB NAS can be purchased for \$250-400; run it in RAID 1 mode; never use mode 5.

Eventually, purchasing a new HDD every year will become accepted digital documentation practice. A generic 1 TB external HDD within an external case, costs about \$150. Next year a 2 TB external HD will be \$150 or less. Try using the Pricewatch website <http://www.pricewatch.com/> to find the best price for a HDD and external case or a complete External HDD.

The standard recommendation for storage of digital files is for three copies to be maintained, with one on a different media and in a different location. Optical storage is not recommended, so data tape storage is the next logical option. Data tape backup on a small scale is not a very attractive option. Done well, it can cost \$3-5000, because single DTL tape drives are expensive and a block of 10 tapes can run \$5-600, two blocks are needed. Most small organizations use multiple HDDs alone. The recommendation for storing one copy in a remote location is to prevent loss in a fire or earthquake. A location 1000 feet to 5-10 miles should be sufficient, but not ideal. Some folks use a safety deposit box at their bank; others use online systems. Both can be time consuming and expensive for large amounts of data. Every day activity shows that storage on the HDD in the originating device and backup on a NAS (RAID mode 1) in the next room works acceptably well. Enterprises with many folks (or organizations) relying on their information should practice the most stringent backup protocols, involving full tape backup at night and storage of last week's tapes in a bank vault.

IT Department

Many conservators are finding that the IT department has become the default curator of their documentation images. The principal reason for this is that the IT department runs the servers that store institutional information. As such, they often approve the purchase of computer equipment thereby forcing image files to be held on their servers. Once on their servers, the IT Dept can dictate the file size and type. In some instances they have dictated that they will only store JPEG image files.

The best way around this problem is for conservators to store the original RAW/DNG or TIFF files on an external drive they purchase in an independent manner, without the knowledge of the IT Dept. Another method is insist that the IT Dept follow established archival procedures that require 3 copies of the file to be held, with one in a location remote from the location of the main backup tapes. This could be the office of the Conservator (or in their home).

The IT Department has a very different set of working criteria than those needed for storing image files. The IT Dept. wants small files so they can be moved around quickly, not clog-up and be served up quickly to the LAN or internet, be it for internal use via an institutional database or for external use by other sites or organizations. The relatively large size of TIFF dSLR image files (15-40 MB) means that they could clog-up a 100 MB/s Ethernet LAN while also taking up a disproportionate amount of precious storage space on [older] servers. There is little relief from this set of parameters, other than fiber networks, modern multi-TB servers and lots of money.

The experience of the IT department is that text files can be compressed with no observable loss of information and that the (heavily compressed) JPEG images seem to look fine on computer screens. They have little experience with, or appreciation for, the numerical RGB values behind the “screen” held in a TIFF file. Showing them Figures 1 and 3 should help and referring them to the research at <http://videopreservation.conservation-us.org/tjv/index.html>. If they don’t listen, then the issue should be raised in professional forums, and should be approached as “their policies” are causing the Conservation Dept to fail to at follow the established standards in AIC’s *The Guide Digital Photography and Conservation Documentation* (2008) [available from [AIC Publications](#)].

Compressing image files destroys information that cannot be restored later. LZW lossless compression of TIFF files decreases size by no more than 30%, and often only 3%, depending of subject matter. This compression rate is of little value when TIFF files are 100 -1000 times the size of a JPEG file. The true value of a TIFF image is the RGB color numbers that are attached to each pixel. While this data cannot be seen on the screen or in a print, it is of immense added value, over the use of film. More information on this topic was presented at the AIC 2005 General Session in a PowerPoint presentation entitled [Getting the Most from a TIFF Image](#); it is available in the [EMG Library](#).

The reason IT Department’s seek to decrease the number and size of files stored on their networks is that they are probably using older equipment. The older equipment can still serve up information at accustomed rates, so there is little need to spend on equipment rather than staff. A simple solution would be to increase the storage capabilities of the older servers. However, increasing storage on older IT equipment is expensive. Modern servers can add a Terabyte (TB) for about \$1000 or less, but on older equipment the cost might be \$100-150K to add one TB; on 10-15 year-old equipment it might run about \$½ -1 M per TB, if the necessary parts could be found.

One solution is to store full-sized TIFF image files locally, while sending a JPEG version to the IT Department, meeting its requirements. This satisfies institutional requirements, saves staff time and forestalls expensive upgrades to older equipment. An automated “routine” for making JPEG images with a specific resolution, pixel-width and degree of compression can be made and saved in Photoshop; see Jeffery Warda’s 2006 PDF on the subject http://cool.conservation-us.org/coolaic/sg/emg/library/pdf/warda/2006-01-warda-automating_tasks.pdf held in the [EMG Library](#).

Negotiating with the IT department to store TIFF image files locally will require sophistication similar to that used to keep fragile artworks from travel or prolonged display. Remember to:

- **offer to make the JPEG yourself**
- **use your budget for local HDD storage**
- **stress the color fidelity of the RGB values**
- **stress that you are saving IT staff time and money**

Photography departments at most institutions will be an all-digital workplace. This required (recent) purchase of all new equipment including cameras, film scanners and computers with more RAM and storage. The size of this expenditure served to focus the IT Dept on the needs of their professional photographers. Those photographers are using the TIFF format for storing and archiving images. In addition, many institutions are adopting institutional storage of large numbers of, large-sized image files in [new] centralized image servers. Often this involves installing a new dedicated Giganet (1000-10,000 MB/s) over wire or optical fiber and a new multi-TB file server or NAS. Optimistic conservators are suggesting that this commitment to digital imaging will, in time, bring an understanding of image technology to all IT Departments. Some national organizations are storing petabytes [1000 TB = 1 PB] of digital data; your IT Dept will be familiar with this and they may even be planning toward those ends. Seven years ago the [Internet Archive](#) was building \$1000 [very cheap] 1-TB blade servers (each using 6 HDDs) ahead of most everyone else, with 250 TB of data storage, larger than most.

Internal and External Hard Drives

Most computer boxes have space for one to two additional internal HDs. One could add two internal SATA, 7200 RPM, 1 TB HDDs (2 TB) with 23 MB buffer for \$300; this makes phenomenal economic sense. Buying large HDDs from the computer-maker vendor will be easier, but will cost you about 100-300% more (think “options” at the auto dealer). There is nothing easier than adding an HDD to an existing computer, on your own.

External HDDs are just a harddrive in a powered external case. Enclosures can have one or more ports; they include FireWire800, USB2, SATA, LAN (RJ45) or a combination. The size of the HDD is also significant; a 2.5” HDD (used in laptops) is far more expensive than the [standard] larger version. They can be called “Pocket HDD” and can powered from the USB or FireWire bus power source (greater battery drain on notebooks, however). The standard sized 3.5” hard drives require external

power supplies. There are numerous online sources to explain the port speeds and advantages; but computers that originally supported USB 1.0 (very slow, 1.25MB/s) and FireWire400 (50 MB/s) cannot support the faster second generation port speeds without the upgrade of an internal PCI card, or PC card (notebook). If you don't know, assume the slower speed and add an extra \$50-100 to your budget. Because FireWire has peer-to-peer architecture it will perform better, especially when multiple drives are attached. USB connectivity is more common, so will be simpler to use when you need to share files with someone else. While transferring a whole HDD has never been easier using FireWire800 (100 MB/s) or USB 2.0 at 480 Mb/s (60 MB/s), moving 250 GB of files will still take hours depending on your computer's internals. If an external HDD should fail to work, check the power transformer because they fail more often than the HDD itself. The first thing to try is to swap out to a new cable [from your backup supply of 2-3 extra cables]. Cable failure is much more common than any other problem. I've also had the chip in an external case fail, while the harddrive was still active; a new case solved the problem. In the past 15 years, I have only had one of 30+ HDDs fail in use.

The most desirable external image storage solution is to use what digital videographers have been using for years: a dedicated external RAID Array running in the RAID 1 or 6 mode. The use of a modern RAID Array will generally get the attention of your IT department because it is similar to their technology, prudent redundant storage and more economical than past incarnations. Former incarnations of RAID Arrays could cost thousands of dollars for a few tens of GBs of storage when using the far more expensive SCSI drives. Today's RAID Arrays (2 to 4 HDDs in prosumer devices) use inexpensive and reliable SATA drives, which start at \$200-250 and run up to \$11-1400 for 8 TB of storage.

Longevity of Hard Drives

Most modern HDs have a rating of 100,000+ hours MTBF, about 11 years of continuous use, with a 5-year warranty (about half MTBF). Most HDs have MTBF data on the label that is physically attached to the harddrive housing; this information is commonly not found online. Modern HDDs use liquid bearings <http://www.ebearing.com/news2001/news187.htm> and other technology to increase lifetime substantially. Many Seagate and Hitachi SATA drives have a predicted life of 300,000 (35 yrs) to 600,000 hours (69 yrs) some MTBF are even higher.

Only 10 years ago HDDs had a real life of 3 years, or less. Today, far few HDD fail and most of the reported failures are from videographers who send HDD to each other and client via FedEx; so shipping plays a role in failure rates. Most of the HDDs today were migrated and retired (as too small) before they failed. In a few years HDDs will have tens-of-TB capacities. This will makes HDD smaller than 1 TB seem impossibly small and therefore inefficient for use in an external enclosure.

It has been shown that cooling HDD does not increase their longevity of HDD, as has been suggested by some experts. External enclosures made of metal with internal fans are uncommon today, except of the more cost versions. I have not seen any consumer-level enclosures, while looking in earnest over the past year, with internal fans.

RAID Array Systems

Consider using a [RAID Array NAS](#) for image storage. RAID 0 uses all drives in the array for storage while RAID 1 strips the same information on both HDD in a 2-drive system. The redundant distribution of the data using mode 1 or 6 will allow reconstitution of the lost data when a drive that failed is replaced.

RAID 5 was the recommend mode several years ago; however, there have been notable RAID 5 failures [see the web server failure at dpreview.com]. RAID mode 6 is the current recommend option, however it is generally not found on consumer or prosumer equipment. I use RAID 1 on my NAS backup systems; a 2 TB NAS hold 1 TB of data. Wikipedia has an excellent entry on RAID <http://en.wikipedia.org/wiki/RAID>.

CD and DVD Optical Storage

There are differing opinions concerning the life expectancy (LE) of optical discs. In 1996 the NML rated the average CD±R at 2.5 years LE, these 10-year-old predictions have proved correct for some of the older CD±R. Joe Iraci (CCI Senior Scientist) rates the **average modern CD±R** at 15 years, but this has yet to be tested by time. An average CD±R is projected by Iraci to have a green dye layer (Phthalocyanine) with silver-colored backing and to have been burned at 48x speed, without verification; it is not a recipe for longevity.

There is no doubt that the CD±R and DVD±R recording process can be made quite reliable with:

- **Outside testing of your system (optical disk testing service bureaus)**

- **Slow write speeds of 2X - 8X**
- **Reliable media such as Kodak or MAM (or their resellers)**
- **Phthalocyanine dye layer**
- **Gold reflective layer**
- **Committed technician**

CD and DVD that are bought already containing music or video are generally pressed using a stamp and have a physical pit which the laser can read. A DVD±R or CD±R has a dye layer in which a pit is formed; the gap is later read by the laser. The manufactured optical disks have a greater longevity because they have a physical bit to read. They will last much longer than their dye-based [±R] counterpart.

Everyone has CD±R that failed within 2-3 years, and others that are older and still remain readable. The latter are on borrowed time. If they are valuable store the contents on a HDD. If you have a collection of writable [±R] optical disks that were used for image storage years ago, migrate them as soon as possible to any available HDD. Do it now.

Both CD±R and DVD±R have much less longevity than predicted by their manufacturers. Kodak Ultima CD±R, Gold-on-Gold and Silver-on-Gold with Phthalocyanine dye layer, were rated by Kodak at 217+ and 100+ years respectively. The NML predicted 50 and 25 years LE, respectively. The Kodak Gold-on-Gold CD±R were discontinued as too expensive to make, but the Ultima Silver-on-Gold are [were] still available.

MAM (formerly Mitsui) is the industry leader. They rate their Gold/Phthalocyanine CD±R at 300+ years <http://www.mam-a.com/technology/quality/longevity.htm> but their actual life is probably no more than 35-50 years. MAM predicts 100+ years for their Gold DVD-R, but the actual longevity is probably about 15 years (three times less than CD±R) because of higher density (6x higher) and lack of a Phthalocyanine dye layer. Many optical media distributors' re-label MAM (OEM) disks as their own brand, making them as good as anything available. Check the AMIA-L listserv archives for details. In Photo Marketing, Ron Kubara (Noritsu Ltd) wrote ***Not all CDRs are Created Equal*** it can be found at http://www.mam-a.com/library/documents/News%20Articles/_Not%20all%20CDRs%20are%20created%20equal.pdf; it has many interesting details on the construction of optical disks.

Hard Drives Storage

Always live, was the best practice for HDD. However, one very knowledgeable magnetic media expert, who serves on many HDD manufacturer technical advisory committees, says that HDD storage on a shelf (not connected to power) is acceptable long term storage, as long as the drive is migrated to a modern drive every 3-5 years and is backed-up on one or more additional HDD. A new device, the [HD Docking Station](#) (\$25) can be used to easily check the viability of the HDD that are not live; it is hot-swappable and fits both size formats. Network administrators recognized that MTBF data predicts that HDDs will be migrated to newer larger HDDs long before they will fail, after 60,000 to 150,000 hours of use.

Additional Digital Storage Information

Readers can find an older version of this essay posted on the EMG website http://cool.conservation-us.org/coolaic/sg/emg/library/pdf/vitale/2005-11-vitale-storage_image_files_long.pdf in the [EMG Library](#). It contains additional CD±R and DVD±R information and an image of the NML life expectancy (LE) predictions before it was pulled from the web 15 years ago. The EMG Library also has a PDF of Joe Iraci's publication on [The Stability of Optical Disc Formats](#) from *Restaurator* (v26, #2, pp 134-150); unfortunately it has no life expectancy (LE) predictions but it does rank the types of optical media from 1 to 12.

Tim Vitale
Paper, Photographs &
Electronic Media Conservator
Digital Imaging & Facsimiles
Film Migration to Digital
Preservation and Imaging Consulting
Preservation Associates
1500 Park Ave.
Suite 132
Emeryville, CA 94608

510-594-8277
510-594-8799 fax
<tjvitale@ix.netcom.com>