How Do You Protect Against Data Loss?

A very important, but often neglected, aspect of digital photography is developing and faithfully using a data backup strategy. Ask yourself, "How much money and effort did I spend to capture my photographic images?" Also consider, "How would I feel if some or all of those images were lost forever?" In addition, "Is there a financial consequence if I lose these images?"

One of the wonderful things about the digital image age is that we can make as many copies of an image as we want, and each one will be a perfect copy, just as sharp and with all the subtle detail of the original. This is true for all digital media and is what scares copyright holders of music, movies, and photography into fits of panic. But that will have to be a topic for later discussion. The good aspect of this ability to make lossless copies is that, as photographers, we can devise protocols to back up our images in order to protect against data loss.

By far, the greatest threat to your data comes from electrical or mechanical failure of the hard drive itself. But, what are the other forces that are trying to deprive us of our artistic endeavors? They are plentiful and varied, including, but not limited to: power surges, brownouts, and outages, fire, floods, earthquakes, tornadoes, hurricanes, tsunamis, equipment failures, terrorism, theft, vandalism, software malfunctions, malicious software, and accidental human error. Whoa, what chance do we have against these forces of man and nature? While we cannot devise a system that is 100% effective, we can develop a strategy that is cost effective and reduces the overall probability of complete data loss to a vanishingly small number. How do we do this?

It is as easy as three, two, one…

Most computer experts now agree that the starting point for developing a comprehensive backup strategy is the 3, 2, 1 rule. What is the 3, 2, 1 rule?

- **"3":** you need to have at least three copies of your digital data
- **"2":** the data should be on at least two different media
- **"1":** at least one copy of the data should be stored off-site

So, let's examine these three rules more closely. First, you need three copies of your data. When you take your images with your camera, you have a copy on the data card. Obviously, the data is not yet backed up. When you get back to base, you copy the data

from the camera compact flash (CF) or secure digital high capacity (SDHC) data card to your computer. You now have two copies, so you have a backup, but have not satisfied the 3, 2, 1 rule, yet. If you erase the data card at this point, you won't even have a single backup. So, once you download your camera data cards to the computer's hard disk, you will want to make additional backups before you erase your data cards.

The "2" rule is harder to meet for a lot of photographers. If you have a small amount of image files, you can certainly make copies of the data onto another media such as CD-ROM, DVD disks, flash memory drives, or copy the files to an Internet "cloud" location. If you are a professional with many terabytes of image data, you may have to stretch the meaning of "different" to distinguish between internal and external hard disk drives. One of the main focuses of different media is because storage media are constantly evolving. If you are diligent to transfer your backup media to keep up with new technology, you will be safe.

The "1" rule takes on different meanings for when you are in the field, "on assignment", and when you are at home in the studio or office. In both instances, the spirit of the rule is "don't put all your eggs in one basket". In the field, this means that all your backups should not be placed in the same bag as you travel. Keep one backup copy on your person, in a pocket. Keep another in your computer satchel. Keep the third copy in another bag, preferably one that is also in your possession at all times.

Once home, the "1" rule means that at least one copy of your data should be in a physically different location: in the "cloud", with a stock agency, in a safe deposit box at the bank, at your vacation home, or with a colleague or family member in a different city or state.

Ask ten photographers how they have implemented a backup strategy and you will probably get ten different answers, but each will probably be equally valid and useful. You can benefit from learning about these different approaches and using the experiences of others to develop a plan that best fits your needs and resources.

What are my backup strategies and protocols?

The concept and procedures for data backup is a fulltime occupation for a huge legion of professionals. One size does not fit all, and the protocol you select should be tailored for your needs and budget. No system is 100% effective and everyone must examine the risk reduction versus cost before choosing a plan for himself or herself. This document is focused on describing the backup protocols that I currently use.

I have two separate backup protocols. One is used while I am in the field, "on assignment", and other is used once I am back in the office. The "on assignment" protocol is used to capture images and back them up as securely as possible such that there is a high probability that the data will safely arrive at my studio/office. Once at the office, my "studio" (primary) backup protocol is used to protect the data from as

many loss factors as possible while not requiring an inordinate amount of my time and keeping my costs reasonable.

Protecting my photographs from loss while "on assignment"

The job of protecting image data from loss must begin in the field while on assignment. Figure 1 is a schematic of my "in the field" backup strategy. It includes redundant means of getting my image data from the camera into my laptop computer and also for making multiple backup copies of the data on external portable disk drives.

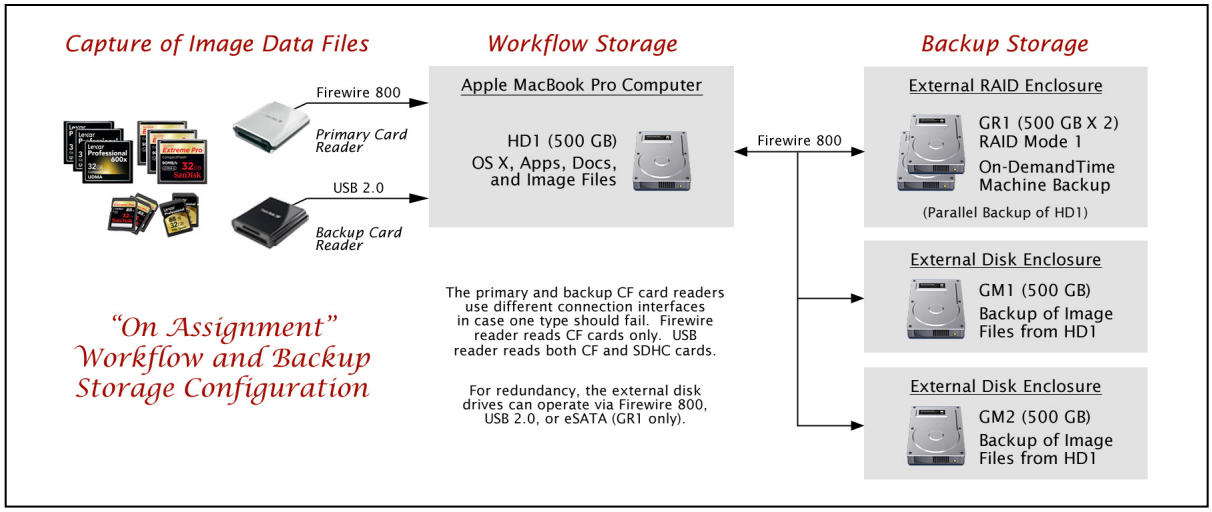


Figure 1. A workflow and backup storage configuration for use while "on assignment" working in the field.

The data protection workflow begins with downloading captured images from the data cards in the cameras one or more times during each day. I read the data cards with an external fast Firewire card reader connected to my laptop computer. I also carry a backup card reader that operates on the USB bus in the event that the Firewire port on the computer is compromised. Although I also carry a small cable that will let me directly connect the camera to the laptop computer, I prefer to use an external card reader for several reasons.

1. First and foremost, an external Firewire card reader is very much faster at reading the data cards than either a USB camera connection or even an external USB card reader.
2. Using an external card reader does not rely on the battery in the camera staying charged during the transfer.
3. I can fit a new blank card into the camera while the "exposed" card is being copied. I like to keep my cameras "ready to go". There is no rule that says an exciting image opportunity can't occur while you are downloading a card.

Immediately after the cards are downloaded to the computer, the next step is to make a backup of the data transferred to the computer onto an external hard disk. On the road,

I use small portable hard disks that are rugged, yet lightweight, and are powered through the connection port of the computer. I prefer external drives that can be connected via the Firewire 800 port of the computer and also have the ability to connect to the USB port in case something goes wrong with the Firewire port.

I find that an external drive with a capacity of 500 GB works very well for expeditions of up to about five weeks in duration. These drives are available from a number of manufacturers including: G-Technology, wiebeTech, LaCie, Western Digital, Seagate, Iomega, and others. I currently use the 500 GB G-Technology G-DRIVE mini disks that feature a rugged aluminum case, Firewire 800 and USB 2.0 connections, and spin at 7200 rpm.

Once an external backup of the image data is made, I will have obtained a reasonable short-term level of protection from data loss. If time permits, a third backup copy should be created right away. In situations with tight time constraints, the third copy can wait for the evening break. My policy is that the data must be transferred to the computer hard disk and at least one other disk drive before I will reformat the camera data card.

I have recently added one more level of backup redundancy to my "on assignment" data protection plan. In addition to the copy of the data on the laptop hard drive and the two external portable hard drives, I make a complete incremental backup copy of the entire laptop computer hard disk to an external disk drive each evening. I make this backup using a program, called Time Machine, that is included with the Apple OS X operating system. I record the backup data on a G-Technology G-RAID mini disk drive. The G-RAID disk drive is powered by the Firewire 800 connection to the laptop computer and has two identical 500 GB disks inside. The unit is operated in RAID (redundant array of independent disks) mode 1, which means that the computer sees just one 500 GB dive, but the external disk unit is actually making two copies of all the data simultaneously. At the end of this document, I provide additional information about RAID and the Time Machine program.

Okay, so I am a bit of a fanatic about data backup. By following my protocol, at the end of each day while I am in the field, I have five copies of the data: computer hard disk, two separate external disks, and two more copies on the RAID disk system.

This workflow does have a weakness. Since backing up my data to the external hard disks requires using the laptop computer, what happens if the laptop computer fails, is lost, or runs out of power while I am on the road? One solution is to haul around a backup computer, but that is often not practicable. As a safety precaution, I try to travel with enough memory cards for my cameras such that I can capture all my images on the expedition without needing to reuse a memory card. Does that remind you of the old "film" days? That data would not have a backup in the field, but at least the trip would not have to be aborted.

Protecting my photographs from loss when back in the studio

My "on assignment" protocol increases my chances of getting home with my captured images. Once home, my "studio" protocol then comes into effect to make sure that I can process my images and protect them from loss by all the natural and man-made forces that conspire to take my data from me.

The first step is to transfer the field data to the workflow files on the hard disk of my desktop computer in the studio. This is the primary copy of the images that will be processed and worked with on a daily basis. Before any processing begins, though, multiple copies must be created. Not shown in the diagram below, copies of the image data files are made to external disk drives. In addition, an on-line Time Machine backup is made of the images as well as all the other files on the computer. The backups are *automatically* created every hour. These incremental backups provide hourly backups for the past 24 hours, daily backups for the past month, and weekly backups indefinitely. I discuss the Time Machine backup utility in more detail at the end of this document.

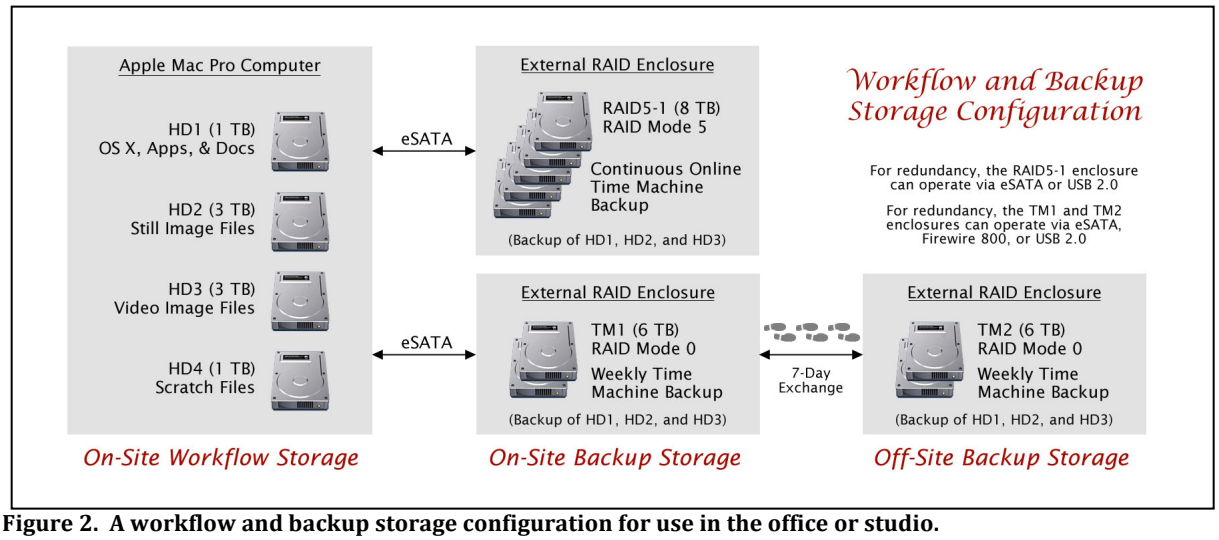


Figure 2. A workflow and backup storage configuration for use in the office or studio.

As part of my backup protocol, I manually create additional Time Machine backups one or more times a week on a separate external disk drive (TM1 and TM2 in Figure 2). This external disk drive is then physically swapped for a similar drive on a weekly basis, with one of these drives taken to and stored off-site.

The disk drives that I currently use for this purpose are made by G-Technology and are marketed as the G-RAID disk drive. Each drive is mounted in a rugged aluminum housing and contains two 3 TB disks and an internal RAID controller operating in mode 0. What this means is that the computer sees the unit as a single 6 TB disk drive. I have included additional information about redundant arrays of independent disks (RAID) at the end of this document. I connect the G-RAID drives to the computer using the eSATA interface, which is a fast interface that is the "e"xternal version of the SATA interface

used by the internal disk drives of the computer. The G-RAID drives also can be connected to the computer using a Firewire 800 or USB 2.0 interface, which provides redundancy and versatility.

The automatic hourly Time Machine disk drive shown in Figure 2 is an external RAID tower of disk drives operating in RAID mode 5. As discussed in more detail at the end of this document, RAID mode 5 links a group of disk drives so that they appear as a single huge-capacity disk drive to the computer. In addition, electronics in the external disk tower records error correction data automatically such that no data is lost even if any one of the disk drives in the tower fails. The RAID tower that I use contains five 2 TB disk drives. The RAID controller uses one-fifth of the disk space for error correction data, leaving 8 TB of usable disk storage for backup data. The tower connects to the computer with an eSATA interface, but can also be connected via the USB 2.0 interface, if necessary.

Is it worth the effort and does it work?

Ask an expert what percentage of disk drives fail, and if they are worth their salt they will reply, "100%". The question is not *if* a disk drive will fail, but rather *when* a disk drive will fail. Given this, it behooves us all to implement a backup protocol.

Unfortunately, I recently had an opportunity to test my backup plan. Fortunately, my plan worked great, resulting in a happy outcome. A rogue application irrecoverably scrambled the primary hard disk on my main workstation computer. Because of my backup protocol, I was able to be back up and running with *NO* data loss within just a few hours. The main hard disk had to be reformatted and the Apple OS X operating system reinstalled. My automatic Time Machine backup was only five minutes old at the time of the system crash and was able to restore all my settings, applications, and data as part of the system rebuild. After the restoration, even my Desktop appeared exactly as it was before the crash.

A small number of manual operations were required to complete the restoration. First, I had to make sure that all the security updates to the operating system were downloaded. I had to re-install the drivers for my eSATA interface cards and for my printers. Two applications, both from Apple, ironically, required that their serial numbers be re-entered. All other applications, data, and settings were automatically restored.

Backup strategies, properly implemented, do work. Don't delay. Develop and implement your backup and recovery protocol immediately – the galactic probability dice are rolling!

Listen Up! This is a RAID…

Modern disk drives are pretty fast, but what if you need one that is even faster? There's a RAID for that. Disk drive capacities are getting bigger and bigger, but what if you need a disk that is even larger? There's a RAID for that. And what if you need protection against a physical hard disk failure? There's a RAID for that, too.

RAID is an acronym for *redundant array of independent disks*. A RAID system consists of several disk drives and dedicated electronics and/or software that link the disk drives in a manner such that they appear to the computer as a single disk drive. There are several ways that have been defined for linking the disk drives together. These different modes have been given numbers, such as RAID 0, RAID 1, RAID 5, and others. Each mode has been developed for a particular set of benefits.

In the description of my backup protocols, above, I make reference to my use of RAID 0, RAID 1, and RAID 5 systems. For the purpose of this document, I will limit my discussion to these modes. There are a number of good information resources on the Internet if you are interested in other RAID modes.

RAID 0 *(block level striping)*: This RAID configuration consists of two or more physical disk drives connected together in a manner that creates a single large-capacity disk that also has the benefit of faster read and write speeds. It takes a finite amount of time to access and read or write a block of data to a physical disk. RAID 0 arrays have special electronic controllers that simultaneously read or write sequential blocks of data on different drives in the array. Since the drives in the array can be reading or writing data at the same time, the speed goes up almost linearly with the number of physical disks in the array.

My use of RAID 0 is just to have very large capacity disk drives.

In itself, a RAID 0 array does not provide any protection against data loss if any of the disks in the array fail. In fact, the chances of the array failing is *N* times that of a single drive, where *N* is the number of physical disks in the array.

RAID 1 *(mirroring)*: This configuration is constructed with one or more *PAIRS* of physical disk drives. Each pair of disks are connected together by electronics or software such that information written to one disk is simultaneously also written to the twin, providing a backup of the data in real time. When data is read from the array, the controller is smart enough to read sequential blocks from the pair of disks, which nearly doubles the read rate.

My use of RAID 1 in my "on assignment" backup protocol is to provide automatic redundancy of the daily Time Machine backups of the disk in the laptop computer.

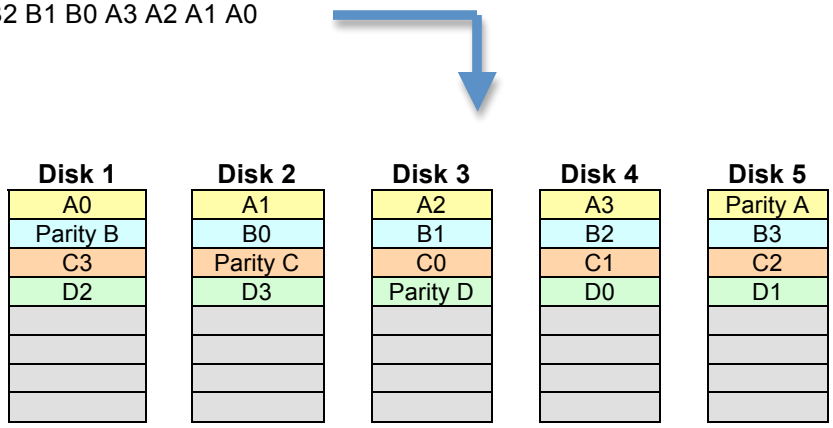
A RAID 1 array provides data protection from the failure of any single disk drive in the array. If a disk in the array fails, the data is still available on the paired drive. The crippled array can continue to be used for reading and writing, but data protection is not provided until the failed disk drive in the array is replaced. When the failed disk is replaced, the RAID controller will automatically copy the data from the surviving drive of the original pair to the new replacement disk. The chances of the other drive in the pair failing before the bad drive is replaced is small if the bad drive is replaced in a timely manner. Note that if the array enclosure suffers a catastrophic physical event (e.g., dropping it in a lake), chances are good that both drives in the pair will fail, so make sure that your third backup copy is in a separate unit. A disadvantage of RAID 1 is that it requires twice as many physical disks than needed to store the backup data.

RAID 5 *(block-level striping with distributed parity)*: A RAID 5 storage system is generally reserved for storage of massive amounts of data with protection against the failure of a drive in the array, and the ability to continue accessing the data even if one of the drives should fail.

A RAID 5 system consists of *three* or more physical disk drives. Unlike a mode 1 array that clones all of the data recorded to the array, a RAID 5 system stores parity information that is used to correct errors on or failure of any one drive in the array. Similar to mode 0, sequential data blocks are stored distributed on the different drives of the array. If the array contains *N* physical disk drives, the equivalent capacity of one drive will be used to store parity data and the capacity of *(N-1)* drives is available for storing data.

My studio backup strategy uses a five-disk RAID 5 array. Here is an example of how a stream of data blocks from the computer might be stored on the five disks in the array:

Stream of data blocks from the computer
D3 D2 D1 D0 C3 C2 C1 C0 B3 B2 B1 B0 A3 A2 A1 A0



The parity information is generated automatically by the electronics in the RAID tower. Notice how the controller spreads the incoming data equally among all the disk drives

in the array and even the parity information is stored on different drives. In this case, each group of five blocks of data (4 of "real" data and one of parity information, e.g., A0, A1, A2, A3, and Parity A) together provide protection against an error in any one of the five pieces of data. The erroneous or missing block can be recreated using the information in the other four pieces of information.

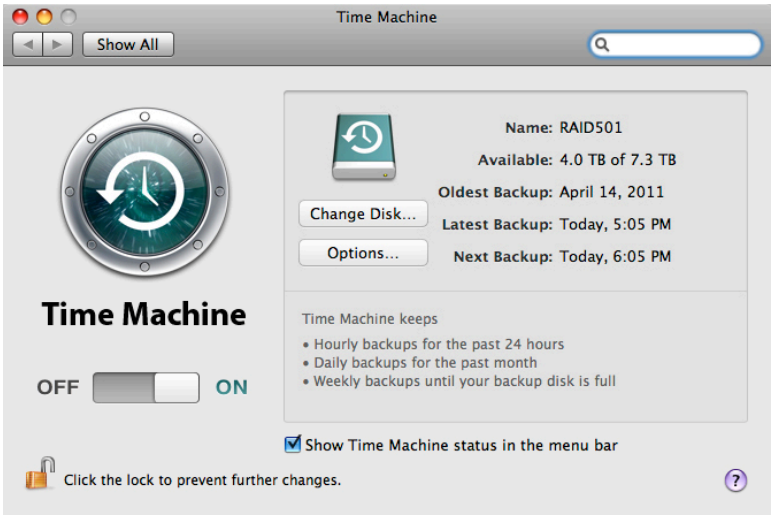
A RAID 5 array protects against the failure of any one drive in the array. After a failure, the array can continue to be used without data loss, but no longer protects against any additional failures until the failed disk drive is replaced. Once a new disk drive is installed, the RAID controller inside the unit will automatically recreate the missing data. Once that process is completed, protection against another single drive failure will be restored. Again, physical damage to the RAID tower might cause failure of multiple drives, so be sure that you have an additional backup stored on another physical medium and preferably stored in another location.

What is Time Machine?

A data backup strategy does you no good if it is not faithfully followed. While we all have good intentions and noble plans to sit down daily, weekly, or monthly, and make the backups of the data stored on our computers, in reality, we are all subject to distractions, overwork, shortage of free time, and so on. We also get complacent as week after week goes by without a loss of data.

What this means is that at least one of your backups should be automatically made on a regular timed schedule. As an Apple Mac computer user, I can use a software utility, Time Machine, which is provided as part of the Mac OS X operating system. A very simple setup screen allows me to select where I want the backup data to be stored, what data I want regularly backed up, and also what information I do not want to back up. When the utility is turned on, it quietly runs in the background.

The first time that Time Machine makes a backup, it may take a long time since it has to make a copy of a lot of data. Later backups are intelligently made by only copying the changes made since the previous backup.



As the name implies, you can effectively travel back in time and recover the backup data as it was on any of the backup dates and times. Time machine saves the hourly backups from

the previous 24-hour period, daily backups for the past month, and weekly backups until the backup disk is full.

There are other third-party companies that make backup software for both Mac and Windows computers. This document focuses on what I have implemented as a backup strategy, so I will leave researching of these other resources to the reader.

Rick Hunter
www.RickHunterImages.com

