

## Archiving digital images – some thoughts

The question is how to store all those digital files as your hard drive (HD) fills up... unlike negatives/prints you can't store them in boxes. First, let's talk about media. There are a couple of options:

1. Another hard drive
2. Optical media (CD or DVD)
3. Tape

That's about it. Remember, whatever media you use, it will fail, sooner or later, just like paper prints or film. So you need a strategy to minimize that risk. You also need boxes to store the DVD's in.

**Another Hard Drive** – advantages, fast on-line access, reasonably priced, you will eventually fill it up and need another one.... Cost, cheap ones similar to DVD/gb, better quality/faster cost more. Durability, good brands, generally good. HD's can be internal, external or "removable", an external drive is useful with a laptop on field trips (or burn to DVD), removable HD's allow you to have several but only use the one you need at any one time. Needless to say, internal drives for a desktop (3.5" format) are the cheapest.

**Optical media** - CD's are getting to be too small if you shoot a lot and/or keep everything "just in case" (I do), DVD's hold more data. There are two kinds of optical media, when you buy a product in a store on optical media it is almost certainly a "stamped" disk. This means the data is physically stamped onto the aluminium platter, which is then encased in plastic. These will last as long as the plastic does. The other type are "burned", this is what you make in your DVD/CD writer (aka "burner"). In these, a laser beam heats and chemically changes a layer of dye, which changes reflectivity/transmission. The data is encoded as dye embedded in plastic, the aluminium platter is just there to reflect light. These are not as durable; life is likely around 3-5 years (depends on brand, don't use cheap "no-name brands if you value your data). Cost is low, around 25-50c per 4.7gb disk. Access and writing time is slow. Always use "write once" (R) DVD's or CD's, not rewriteable (RW) disks, they do not last as long.

**Tape** – no longer a real option for archiving although I use tape for routine backup (different issues with different criteria). Access is very slow (sequential reads, not random access like HD's or DVD's). Cost is high for the hardware and media is now more costly than DVD's/gb, but it's reusable. Depending on the tape type, storage per tape is much larger than DVD but well below HD's.

To provide for data integrity, here is what I do when copying files to DVD:

1. Make 2 copies, if one fails the chances are the other will be readable (see above, the media will fail).
2. Always verify the data, this means reading what was just burnt and comparing it to what's on the HD (this is done by the DVD software if you tell it to). True verification is a bit-by-bit comparison, "quick" verify is not and is useless.
3. I also create a "Check-sum" for each file I'm copying and include that file on the DVD. I use EF Software's "Checksum Manager" which I find works well. I then run the checksum file against the DVD. This reads the file, recomputes the checksum, and if they match, the file is good.
4. Test read a few files off the DVD (usually 1 per directory).

With this process I'm pretty confident that I can retrieve files off the DVD for several years. After that time, it's time to copy them all to new media....

## Other issues

A **good cataloguing system** that allows you to search for files by multiple parameters is essential. To help this I have a few rules that I follow:

1. Always keep the camera assigned serial number as part of the file name (set it to increment, not reset, when you change cards).
2. For “good” images I then add a descriptive “keyword” or two to the file name (e.g. “Kingfisher”, “Guitar” etc.) that describes the main subject.
3. I try and put images for related subjects in the same directory (typically a day shooting wildlife, or studio session). The directory is then named with the date (in YYYYMMDD format so it sorts correctly) and the location.
4. Include the camera body type in the filename if you use multiple bodies (or buy a new camera).

So, in my system a file that has the name D:\Camera\Img\20071212 Esquimalt\D20-16120 Kingfisher.NEF was shot on *Dec 12 2007* at *Esquimalt Lagoon*, using a *D200* (D20-), is the original image (*NEF*), and is a “keeper” of a *kingfisher*. Derivatives of it would have a name like D20-16120 DXO Kingfisher.dng, this is the same image after processing through DXO, and is saved in DNG format. The next would be ... DXO Kingfisher.psd, this is my “working” corrected file ready for pre-print processing. I do not keep “printable” files with sharpening as they are too large and typically would need redoing for use in different media (large print, slide, web use etc). If I prep a file but don’t print it, it gets “Print” added to its name and my backup software ignores it. If I’m looking for other “Kingfisher” files I know that they are likely in the same image number sequence (i.e. the same directory on the same disk). If you don’t retain the image number... good luck finding them.

Each DVD pair gets a unique volume name, I use “Photo nnnn”. This is recorded on the disk data by the software as well as used on the label.

A good cataloguing program will be able to search and sort by disk volume, path, filename, date (created, modified) and parts of the filename (and a number of other parameters). It will also have the ability to store keywords in its database (but remember that this is proprietary, if you change software you will likely lose it).

**If one of your disks fails to read** ... remember that you have a backup (you did do this didn’t you?). So use that, and immediately make a copy of it. If disaster strikes and the backup also fails, there are a couple of things you can do.

- The major flash media (CF, SD cards etc) manufacturers produce “Rescue” software that can retrieve files from some problem disks. A DVD is similar to the camera media in that the files are stored in the same encoding method. Similar software is also available from other sources.
- Send the media to a data recovery lab (you need to be really desperate, it’s not cheap!)

The recovery software works by reading the physical media without using the “index” – aka FAT, that tells the computer OS where to find the data on the media. This works well if it’s the FAT that is corrupt/unreadable, you should get most files back (if you are doing this on flash media or a HD, you will also get a pile of “deleted” data back as well). If the corruption is elsewhere on the media, the software will read up to that point, then try and skip the bad area and continue with the remaining good area. That way you will get all the “easily recoverable” data back. [Note: “deleted” data is not really gone, the OS just changes the first bit of the filename to a reserved character so that it is ignored by the system and treated as “free space” until reassigned to another file and overwritten. To really delete it you have to “wipe” that area of the disk that involves overwriting it many times with random data.]