Patrick H. Corrigan

Data Protection for Photographers

A Guide to Storing and Protecting Your Valuable Digital Assets

rockynook

Patrick Corrigan (www.dpworkflow.com)

Editor: Joan Dixon
Project Editor: Maggie Yates
Copyeditor: Maggie Yates
Layout: Cora Banek
Cover Design: Helmut Kraus, www.exclam.de

Printer: Everbest Printing Co. Ltd through Four Colour Print Group, Louisveille, Kentucky
Printed in China

ISBN 978-1-937538-22-4
1st Edition 2014

Rocky Nook Inc.
802 East Cota St., 3rd Floor
Santa Barbara, CA 93103
www.rockynook.com

Many of the designations in this book used by manufacturers and sellers to distinguish their products are claimed as trademarks of their respective companies. Where those designations appear in this book, and Rocky Nook was aware of a trademark claim, the designations have been printed in caps or initial caps. All product names and services identified throughout this book are used in editorial fashion only and for the benefit of such companies with no intention of infringement of the trademark. They are not intended to convey endorsement or other affiliation with this book.

While reasonable care has been exercised in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein or from the use of the discs or programs that may accompany it.

Chapter 6

Backup and Archiving

Backup and *archiving* both refer to storing copies of data. The difference between the two is their purpose. Backup refers to copying and storing data for use in case of emergency. Archiving is data stored for long-term retention.

When you make a backup, you must consider all of the factors that can affect the use of your data, including accidental file deletion, system failures, and natural disasters. You must not only consider the data, but everything that must be done to get you up and running again. There are four primary purposes for backing up files:

- **Restoring deleted, overwritten, or damaged files:** This is by far the most common reason for restoring from backup. However, if the most recent backup was automatic, it may still be missing the deleted file or have the damaged file. Because of this, the ability to maintain multiple versions of files is important. This can be done by either using a backup media rotation scheme that maintains multiple backups or by using software that can maintain multiple file versions. The traditional method, used for years with tape backup systems, is to maintain multiple sets of backup media. A newer approach, which started with disk-based backup systems, is software-base versioning, in which older file versions are maintained even though newer versions are added to the backup.
- **Restoring after a system failure:** In this case, it is desirable to be able to do a disk image restore, or use a clone disk to replace the original system disk, giving you the ability to start the system without having to do time-consuming data restoration. These approaches (discussed later in this chapter) allow you to restart a system in the exact state it was in at the time of its last backup. This works well in the case of a hard disk failure. This does not always work as well when you are replacing or upgrading a system, since a new system usually has different hardware or a different operating system version from the old system. Fortunately, many (but not all) of the current backup software programs that do imaging can restore to different hardware. Also, many of the software programs that do imaging or cloning allow you to restore specific files and folders.
- **Restoring after a disaster:** In this case, not only is the original system that was backed up likely not available, the building in which it was located might be gone or so severely damaged that it is not available. For these situations you will usually have to rely on backups stored off-site.

- **Archiving:** Requirements for archiving vary from industry to industry, but in nearly all organizations there is a need to archive at least some data. Archiving is often done separately from backup since not everything on a particular computer or disk system may need to be archived, and there may be specific legal requirements for long-term storage of certain kinds of data. Professional photographers may have large libraries of images that are no longer active. Keeping and managing those images on primary storage systems and backing them up regularly is time-consuming and expensive. It is often much better to move inactive files to a separate archive on an off-line disk, optical media, or magnetic tape.

In the case of both backup and archiving, you should maintain multiple copies of data. When you are making decisions about the policies, procedures, and technologies that you will use, keep the purposes for copying and storing that data in mind. Of course, all of this must be done in a suitable time frame. If you make your living as a photographer, you might be able to survive a few days of downtime, but more than that could be disastrous.

What to Back Up

Your data, including your images and business data (customer files, accounting data, contact lists, critical email files, etc.), should be first and foremost in your mind. This is the data you can’t replace. Custom software configuration files, templates, and other data that customizes software should also be backed up. This includes items such as Photoshop and Lightroom templates, and browser bookmarks. This takes some care, because programs and operating systems all bury this data in different locations. Lastly, even though you can reinstall programs and operating systems from original media (assuming you still have such media), how much time will it take? Ideally, you should have the ability to either restore this data quickly and easily, or be able to run from a clone disk if your primary hard disk has crashed. Accomplishing all of the above may require more than one backup approach and more than one type of software. In addition, you have a large variety of backup technologies from which to choose.

Creating a Data Protection Plan

Instead of concentrating on backup alone, I suggest creating a comprehensive data protection plan. A data protection plan is an outline of your data protection goals. It can be formal or informal, but should include the following:

- **Minimization of downtime:** You need to consider the consequences of short-term and long-term downtime. Some downtime is inevitable. If nothing else, most systems must be shut down for occasional maintenance. You need to determine the maximum unplanned downtime you can afford, and then weigh that against the costs of implementing the systems and services necessary to achieve your desired level of uptime.

- **Access to data in an emergency:** Data is backed up so that it is available in an emergency. This could be a disk failure, data corruption, accidental data deletion, or a number of other causes. The backup data should be available and you should have the means to restore or access it.
- **Long-term storage of archived data:** Make sure that your data protection plan provides for archival storage of data that must be kept for legal, financial, or business reasons. Plan to migrate archived data to new media every five to seven years.
- **Verification and testing of backups and archives:** Most file and folder backup programs provide an optional verification process to confirm that the data written to the backup medium is the same as the data on disk. LTO tape drives do an immediate verification after write, checking and correcting for possible tape errors. Even with verification, it is still a good idea to manually check backup media to make sure your data is readable.
- **Moving unused data from spinning media to static media for archiving:** This practice frees up space being used to store inactive data.

An effective data protection plan should include a method to quickly restore a system in case of emergency, and a method to restore individual files and folders, including previous versions. It should include local and off-site storage of backup and archived data.

In an ideal system, you can restore data to any previous point in time or file version. You can perform a *bare-metal restore*, which quickly and easily restores a complete system directly from backup, including system files and data, to the same or different computer hardware. Also, you can restore this data locally or in a remote location. This is what backup system vendors often promise; unfortunately, reality sometimes fails to live up to this ideal. Good policies, procedures, and appropriate backup hardware and software can get close, but always be prepared for a few hiccups.

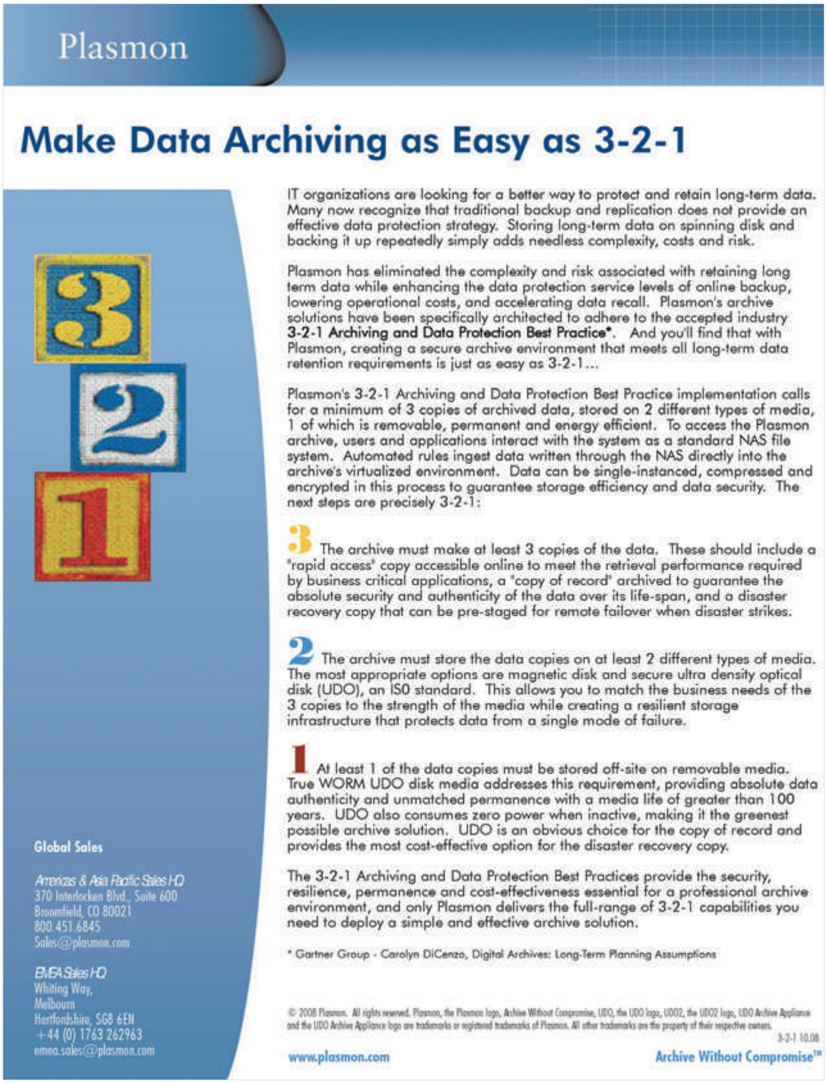
The 3-2-1 Approach

3-2-1 is a simple backup strategy: **Maintain at least three copies of data on at least two different types of media in multiple locations, with at least one copy maintained off-site for recovery in case of a disaster.**

I recommend this simple strategy, but it’s just a general outline; you still need to fill in the details yourself. When it comes to data protection, one size does not fit all.

***The Origins of 3-2-1:** The name “3-2-1” was originally a marketing slogan created by Patrick Dowling and Steve Tongish while they were working at Plasmon. (Plasmon has since declared bankruptcy and Allied Storage Technologies acquired its assets.) The slogan was developed with market intelligence provided by analyst Carolyn Dicenzo, Research Vice President for Gartner, Inc., who supported Plasmon in its 2007 efforts to popularize the concept. “An archive strategy to meet today’s demanding compliance standards is a business imperative. IT executives should strive to separate fixed content data from production data and move static data to an active archive,” said DiCenzo. “A best practice for securing long-term data archives is to maintain three copies of data on at least two different types of media in multiple locations, mitigating the risk of media failure. At least one copy of the media should be removable and be maintained off-site for disaster recovery purposes.”*

After Plasmon went into receivership, Tongish moved to QStar Technologies, where he and Dowling reused the concept of 3-2-1. QStar then produced marketing collateral based on 3-2-1, and other companies adopted it, as well. It has since become firmly established as a basic principle of data backup and archiving.



Plasmon

Make Data Archiving as Easy as 3-2-1

IT organizations are looking for a better way to protect and retain long-term data. Many now recognize that traditional backup and replication does not provide an effective data protection strategy. Storing long-term data on spinning disk and backing it up repeatedly simply adds needless complexity, costs and risk.

Plasmon has eliminated the complexity and risk associated with retaining long term data while enhancing the data protection service levels of online backup, lowering operational costs, and accelerating data recall. Plasmon's archive solutions have been specifically architected to adhere to the accepted industry **3-2-1 Archiving and Data Protection Best Practice***. And you'll find that with Plasmon, creating a secure archive environment that meets all long-term data retention requirements is just as easy as 3-2-1...

Plasmon's 3-2-1 Archiving and Data Protection Best Practice implementation calls for a minimum of 3 copies of archived data, stored on 2 different types of media, 1 of which is removable, permanent and energy efficient. To access the Plasmon archive, users and applications interact with the system as a standard NAS file system. Automated rules ingest data written through the NAS directly into the archive's virtualized environment. Data can be single-instanced, compressed and encrypted in this process to guarantee storage efficiency and data security. The next steps are precisely 3-2-1:

3 The archive must make at least 3 copies of the data. These should include a "rapid access" copy accessible online to meet the retrieval performance required by business critical applications, a "copy of record" archived to guarantee the absolute security and authenticity of the data over its life-span, and a disaster recovery copy that can be pre-staged for remote failover when disaster strikes.

2 The archive must store the data copies on at least 2 different types of media. The most appropriate options are magnetic disk and secure ultra density optical disk (UDO), an ISO standard. This allows you to match the business needs of the 3 copies to the strength of the media while creating a resilient storage infrastructure that protects data from a single mode of failure.

1 At least 1 of the data copies must be stored off-site on removable media. True WORM UDO disk media addresses this requirement, providing absolute data authenticity and unmatched permanence with a media life of greater than 100 years. UDO also consumes zero power when inactive, making it the greenest possible archive solution. UDO is an obvious choice for the copy of record and provides the most cost-effective option for the disaster recovery copy.

The 3-2-1 Archiving and Data Protection Best Practices provide the security, resilience, permanence and cost-effectiveness essential for a professional archive environment, and only Plasmon delivers the full-range of 3-2-1 capabilities you need to deploy a simple and effective archive solution.

* Gartner Group - Carolyn DiCenzo, Digital Archives: Long-Term Planning Assumptions

Global Sales

Americas & Asia Pacific Sales HQ
370 Interlocken Blvd., Suite 600
Broomfield, CO 80021
800.451.6845
Sales@plasmon.com

EMEA Sales HQ
Whiting Way,
Melbourn
Hertfordshire, SG8 6EN
+44 (0) 1763 262963
emea.sales@plasmon.com

© 2008 Plasmon. All rights reserved. Plasmon, the Plasmon logo, Archive Without Compromise, UDO, the UDO logo, UDO2, the UDO2 logo, UDO Archive Appliance and the UDO Archive Appliance logo are trademarks or registered trademarks of Plasmon. All other trademarks are the property of their respective owners.

3-2-1 10.08

www.plasmon.com

Archive Without Compromise™

Original Plasmon ad promoting their product with the 3-2-1 system

Backup Hardware Technologies

You have a choice of hardware technologies to use for backup, and they all have their pros and cons. If you follow the 3-2-1 approach, you will use at least two technologies.

External or Removable Disk

External and removable disks have the advantage of reasonably low cost per GB combined with relative ease-of-use.

External Disks

Plugging in an external disk using USB 3.0, eSATA, or Firewire 800 is simple. These disks are sold in stores like Costco, so the prices are often lower than equivalent internal disks. The disk is in an enclosure, so it has some protection from shock. 2.5” external disks usually take power from the computer through the data connection, but 3.5” disks require an external power supply. There’s no standardization, so swapping disks may also mean swapping power supplies and power cords.

Removable Hard Disks

Removable hard disks have similar advantages and disadvantages as external disks. Removable disks are usually bare (internal) hard disks that are inserted into an external docking station or a carrier installed in a 5¼” device bay in a computer. The disk carriers use standard SATA data and power connections, and the external docking stations use eSATA or USB 2.0 or 3.0. Some carriers and docking stations only accept 3.5” disks, but some accept both 2.5” and 3.5” disks. If you are using a disk carrier connected to an internal SATA port, it is a good idea to set the SATA port for *Hot Plug* in your computer’s BIOS setup. This allows you to safely install and remove disks while the system is running. If you are using bare disks, it is a good idea to use a case or protective device when transporting and storing them.

Issues with External and Removable Disks

The advantage of external and removable disks is that they can stay connected and powered on or they can be disconnected and stored elsewhere. If they remain connected, they are subject to many of the same potential problems that can affect your primary storage, such as power-related issues, accidental file deletion or corruption, viruses and malware, and normal disk failures. Disks that are disconnected and stored can degrade over time. Disk manufacturers do not quote a shelf life for disconnected hard disks, and most have an average life of five years for disks in normal use.

In addition, the life of your disk-based backups is limited by the availability of the technology needed to connect them to a computer. For example, the ST-506/412 interface created by Seagate in 1980 was the major disk interface standard into the 1990s. Today, if you had a working ST-506/412 drive, you would be hard-pressed to find a way to connect it to a modern computer. This is one reason to periodically refresh your long-term archives to newer media. Another issue is not with the technology itself, but how people use it. Many people back up or copy data to a single disk and think they are protected. Even worse, they continually copy new data over older data. Remember 3-2-1!

Disk Cartridges

There are several disk cartridge systems on the market, but the only one with multiple vendor support is the RDX cartridge system that was developed by ProStor Systems Inc. in 2004. Numerous companies have licensed the technology, and the product is available under many brand names, including Tandberg Data, HP, Imation, and Quantum. The cartridge incorporates a 2.5" hard disk in a hermetically sealed plastic cartridge. Internal carriers and external docking stations are available for RDX. The major issue with RDX and other cartridge systems is cost: disk cartridges typically cost two-to-three times that of bare or external disks.

RDX vendors claim the cartridges have an archival lifetime of thirty years, and can sustain a 1-meter drop on a concrete floor. This claim is based on a single accelerated-aging test that used high heat and humidity (conditions to which disks are not usually subjected) to test the cartridges. The test did not determine magnetic signal fade over time, something that affects all magnetic media to one degree or another. You can see the study here:

dataarchivecorp.com/pdf/prostor/RDX%20Removable%20Disk%20Archivability%20Study.pdf

Optical Disc

For backup purposes, CDs are no longer viable for most photographers, due to their limited capacity. DVDs are also less viable than they were a couple of years ago, because the size of digital images has grown dramatically. Less than two years ago, for example, my RAW images were about 10 MB in size; now they're about 25 MB. A single-layer DVD can hold fewer than 200 images, while a double-layer DVD would hold fewer than 350.

Even thought the cost of Blu-ray discs has dropped substantially, in terms of capacity, they make less sense for backup than they once did. I currently use 32 GB and 64 GB memory cards. A single-layer BD holds 25 GB and a double-layer holds 50 GB, so backing up a single 64 GB card would require at least two discs. I could use BDXL, but at $40 and higher per disc, I can't justify the cost.

There is also the issue of life span (see chapter 2). CD-Info (cd-info.com) claims a shelf life of 50-200 years for recorded CD-Rs, 30-100 years for recorded DVD-Rs, and

30-200 years for recorded BD-R and BD-E. On the other hand, a document posted on the U.S. National Archives website (http://www.archives.gov/records-mgmt/initiatives/temp-opmedia-faq.html) says, "CD/DVD experiential life expectancy is 2 to 5 years even though published life expectancies are often cited as 10 years, 25 years, or longer. However, a variety of factors...may result in a much shorter life span for CDs/DVDs." The difference between pre-recorded music and video CDs and DVDs that you buy in the store and writable CDs and DVDs is that the pre-recorded discs are pressed, while the writable discs are burned.

If you still plan to use optical media for backup, I would suggest reading the National Archives document referenced above. Also, make sure optical media is not your only backup media.

NAS or SAN Disk Array

The big advantage of NAS and SAN devices is capacity. In addition, some arrays allow you to expand capacity by adding disks or replacing low-capacity disks with larger-capacity disks. These devices can give you enough capacity to store backups of your data, and images of your boot disk or partition for quick recovery in an emergency. If your budget allows, this is a good option for the primary backup system in your 3-2-1 strategy.

If you have more than one computer on your network, a NAS probably makes more sense than a SAN, since it allows you to back up all your computers to the same device. Many of today's devices function as a NAS or a SAN, and some allow both functions concurrently, although the two functions cannot access the same data. Bear in mind, however, that NAS and SAN devices are computers, and subject to the same issues as other computers. Any component in the device can fail, including multiple hard disks. The devices typically remain powered-on all or most of the time, so they consume a significant amount of power. If you use a NAS or SAN device for backup, you should also consider using some form of off-line storage, such as removable disks or magnetic tape, as a second form of backup.

Magnetic Tape

Although disk-based backup has become popular in recent years, the use of magnetic tape continues to grow as well. There are several reasons for this:

- **Tape is reliable.** Current tape technology, such as LTO Ultrium, has proven to be very reliable. When properly handled and stored, LTO storage life is estimated to be about thirty years. Based on bit error rates for tape and sector error rates for desktop SATA disks, LTO is about 100 times as reliable as disk.

It is prudent to migrate data from tape or any other backup or archive media every five to eight years.

- **Tape is portable.** Magnetic tape is easy to transport or ship. Current tape technologies are very resistant to shock.
- **Backing up to tape is faster than backing up to disk in most circumstances.** In the past, backing up to tape was slow because it had to be fed at constant speed. Modern LTO drives use variable speed and buffering to alleviate this problem.
- **Tape is cost-effective.** Although the per-terabyte cost of hard disks continues to drop, so does the cost of tape. For example, at the time of this writing, the street cost of an LTO-5 tape cartridge with a native (uncompressed) capacity of 1.5 TB is $33-40. The cost of a 1 TB, 3.5” desktop-grade hard disk is $50-80. There is, of course, an initial investment in the tape drive and SAS host adapter.
- **Tape is safer.** Because tape is taken offline, it is not susceptible to the same system and software glitches that can affect disk-based backup systems. Also, since tape doesn’t use the same file systems as disks, it is less susceptible to the malware that can corrupt disk systems.

All major video production studios use tape because of its cost-effectiveness, reliability, and suitability for long-term storage. Even online storage services use tape. On February 28, 2013, a software update effectively deleted all on-line copies of about 0.02% of Google Gmail data. Fortunately, Google backed up that data to tape. Google’s blog entry about the incident is here: gmailblog.blogspot.com/2011/02/gmail-back-soon-for-everyone.html. A third-party blog post is here: blog.bandl.com/2011/03/07/why-the-google-incident-proves-relevance-of-tape-storage.

Tape does have some disadvantages:

- **Tape is a linear medium.** It does not lend itself to deleting or changing individual files. This is the case even with LTO’s Linear Tape File System (LTFS). When you “delete” a file from LTFS, the space used is marked as deleted, but it is not released for reuse. Even though tape records on multiple parallel tracks (LTO-5, for example, writes to 1280 parallel tracks, writing to 16 tracks each pass), restoring a single file is slower then restoring from disk because the initial seek time is longer. Seek time for LTO tape is in the tens of seconds, while seek time for a disk is in milliseconds. Restoring a large block of contingent data from tape should not be slower than restoring from disk (other than the initial seek time) since restore time is limited by the write performance of the target disk. In other words, the limiting factor is how fast the disk being written to can ingest the data, regardless of whether it comes from tape or another disk.
- **Tape storage is static.** Disk backups can be dynamic, with constant, continuous updates. All updates and changes to data on tape must be added to the end of the existing data. This makes tape unusable by some backup programs that maintain multiple file versions. Of course, the backups created by those programs can also be backed up or archived to tape for an additional level of safety.

- **Current LTO tape drives require a SAS connection.** For a desktop system, this means adding a PCIe SAS adapter or, for Macs, a Thunderbolt-to-SAS adapter. Newer Mac laptops have Thunderbolt, but Thunderbolt-to-SAS adapters are expensive. Most other laptops have no provision for adding a SAS adapter.
- **There is an initial cost.** You must buy a tape drive and SAS adapter.
- **Tape life is usage-sensitive.** Tapes wear with usage. For example, an LTO-5 tape is estimated to be good for about 16,000 end-to-end passes.
- **Tape is "different."** Dealing with tape is different from dealing with disk. Unless you are using LTFS, tape can only be accessed through your backup software.

If you have more than 1 TB of data, and sometimes even if you have less, I recommend considering LTO-5 or LTO-6 tape as a part of your backup and archive system. LTO drives will write to and read from the previous generation (i.e., LTO-6 will write to LTO-5). Even if LTO-5 meets your current needs, consider future-proofing by buying an LTO-6 drive and using LTO-5 tapes. As your library size grows, you can always move to the larger-capacity LTO-6 tapes.

Online (Cloud) Backup

Online backup can give you additional protection against deletion or corruption of critical data files, but it should never be your only backup. Use an online backup service that allows you to retain multiple versions of backed up files and set the number of file versions to retain, as well as the file retention period. On-line backup should be used for critical data files that cannot be easily recreated. Usually software and system files do not need to be included in online backup (assuming you have those included on other backups), but you may want to include configuration files and templates that are often intermixed with application software.

Security is a major concern when using online backup. You have no way of knowing who has access to your data, regardless of service provider promises. Encryption makes sense for online backup, but be aware that encryption can be a double-edged sword—if you lose the security keys, your data can remain locked up forever.

Be aware that events outside your control can affect your access to data. Cloud services do have outages and lose data. Cloud services even shut down (or are shut down by government agencies) with little or no notice, taking your data with them. Protect yourself. Never make cloud backup your only backup.

Private Cloud

Another approach to cloud storage is creating your own *private cloud*, which is private storage accessed from the Internet. Many NAS device vendors include this capability, although they typically have their own name for it. LenovoEMC (formerly Iomega) calls theirs *Personal Cloud*, while Netgear calls theirs *ReadyCloud*. Although there are differences in the services from each vendor, the basics of access are the same: you register the unit with the vendor's online service, then log-in to the service, authenticate, and the service gives you a direct connection to your remote NAS device. There is a software component to install on your computer, which allows you to set a path or map a drive letter to the device. You can typically access your system from a mobile device or web browser as well.

This approach has a couple of advantages over a cloud service. First, you are in control: you know where your data is physically stored. You are not subject to the whims of a cloud service provider, such as abrupt changes in service plans and costs. You know exactly what your costs are, because you paid them when you bought your NAS device.

There are some disadvantages, as well. Depending on the system, you may have to make firewall and router modifications to allow communication to and from your remote device. You also need a place to put the device. It could be anywhere with a good Internet connection and good power, but remember that once it's in a remote location, you do not have immediate hands-on access and control, so you must trust whoever is at that remote site. You also have fixed costs—you have to pay for the hardware, and typically, more storage than you immediately need. With cloud storage, you pay as you go.

Your remote location should be far enough away to protect your data from a site disaster but close enough to physically access if you need to. If you are concerned about natural disasters, find a storage site that is unlikely to be damaged in such events.

Backup Methods

Not too many years ago, you had two options for backup—file and folder backup, and image backup. Today there are many backup options, but not all of them work effectively (or cost-effectively) in small offices, or for all operating systems. This book will concentrate on backup systems that can be effective in small offices. There is a whole class of backup software and systems targeted toward the enterprise (large organization) market, but that is beyond the scope of this book.

File and Folder Backup

This method backs up files and folders and their related metadata, such as security and access control information. Even if you back up the files required by your operating system, this approach, by itself, will not let you build a working system after a crash. You will first need to reinstall the computer's operating system and, most likely, all of your programs. In other words, this form of backup is great for restoring your data, but not so great for restoring a complete system.

File Synchronization

File synchronization refers to the periodic or continuous copying of files and directories from a source location to one or more destination locations in order to maintain duplicate file sets. This technique is often used to make sure the most recent versions of files are available elsewhere if a primary system fails. When implemented with a versioning system, this approach can maintain multiple revisions of files. File synchronization is often used between systems within an organization to make sure data is quickly available in case of a site-related disaster. File synchronization is often used in addition to traditional backup systems since it can provide immediate access to data. Most approaches to file synchronization are unidirectional, meaning they synchronize in one direction only.

Bidirectional or multi-directional approaches also exist, but they are much more complex to implement, and often require manual intervention to avoid version conflicts. When updating files that have previously been replicated, some programs re-replicate entire files, while some use a technique called *delta encoding* to only replicate file changes. Delta encoding can significantly reduce disk space usage, network traffic, and replication time.

Image Backup

This method creates an image of a disk or disk partition that can be used to restore the system to the state it was in at the time the image was created. With an image backup you will typically boot your system with an optical disc (CD or DVD) or a USB flash drive, and install the image on the disk. This process will overwrite anything that is on the disk; if your disk has multiple partitions, and you have a partition image, it will overwrite anything in that partition. The image file will be smaller than the disk or partition. This is because the file only contains data, and does not include the unused space on the disk or partition. In addition, many imaging programs use compression to further reduce the image file size. Depending on the software used to create the image, individual files and folders can usually be extracted from the image and restored.

The biggest advantage of image backup is that it allows you to quickly restore your system. However, you must back up the entire partition or disk, and you cannot selectively back up files and folders. Doing this on a daily basis and keeping multiple versions can take up a lot of disk space. Some imaging software, however, will create incremental or differential backups which only save changes. Compared to storing multiple full images from backup, this saves disk space. These files must be used with the full backup to restore a system to the most current version. Another problem with image backup is that depending on the imaging software, you may or may not be able to restore your image to different hardware.

Cloning

Cloning a disk means making a bootable copy. If your primary disk fails, you can immediately boot the system from the clone. The big advantage to cloning is that you can be up and running immediately after a disk failure. Also, since the clone is a copy of a disk, if it is mounted as an additional disk (not the boot disk) all files are accessible with programs and standard file management utilities. The big disadvantage is that the clone disk, in most cases, should have at least as much capacity as the cloned disk, which means maintaining multiple versions will require multiple disks. Unlike an image, which is a file and can be stored on disks holding other data, each clone is its own disk. As with imaging software, a clone may only work with the original or an identical computer.

Snapshots

A snapshot captures the state of a system, volume, or directory at a specific point in time. Disk images and clones are actually forms of snapshots, but there are other forms as well. Snapshotting can be a function of a file system or backup software. Snapshots created by the file system are usually stored on the same volume as the files and directories being snapshotted. Snapshots created by backup software, however, are typically stored on separate backup media. For example, the BTRFS file system used on some Linux systems and NETGEAR ReadyNAS devices can create snapshots of directories and then store those snapshots on the same volume. Apple's Time Machine stores snapshots on an external hard disk.

Snapshotting techniques vary greatly in how they capture and store data. For example, when creating an initial snapshot, BTRFS does not actually copy any data, it just sets pointers to the original data. Subsequent snapshots then only record the changes and store those parts (deltas) of files that have changed. Time Machine, on the other hand, copies all the data being tracked to the Time Machine backup drive. On subsequent snapshots, it saves entire changed files to the backup drive. In both cases, entire data sets or specific files and folders can be restored to a particular point in time.

> *Pointers are additional references to existing objects or files in the file system. For example, if BTRFS creates an initial snapshot of a particular directory, it doesn't make a copy; it just references the original so the initial snapshot does not take additional disk space. Only subsequent changes take disk space.*

This type of snapshot makes it easy to recover data in a previous state. Even if you delete the original directory, the data that is tracked by the snapshot remains, allowing you to restore the saved version of the entire directory. Of course, if you lose the whole volume, you lose the snapshots as well, since they must be created on the same volume. However, it's a quick and effective way to recover deleted or overwritten data.

Full and Incremental Backup

A full backup will back up everything, while an incremental backup will only back up data that has changed since the last full backup. The full/incremental approach has long been used for file and folder backup, and more recently is being used with images and clones as well. The specific mechanisms vary between different backup programs, but the approach allows data to be restored to a specific point in time, or allows specific file versions to be restored.

Continuous, Near-continuous Backup

When data is written to disk, a continuous data protection system saves that new or updated data to a backup system. A near-continuous data protection system will capture changed data every few seconds, or at pre-defined intervals instead of immediately upon disk write. For most purposes, the two approaches are the same—data can be restored from nearly any point in time. Both approaches can have some effect on system performance, and both generally consume more backup media space than more traditional approaches. In many cases, continuous and near-continuous backups function like full and incremental backups, allowing the restoration of specific files or restoration to a specific point in time. The difference is that full/incremental is typically done on a manual or scheduled basis, while continuous and near-continuous backup monitors the system for changes and performs the backup as needed.

Versioning

Many modern backup systems give you the ability to restore previous versions of data, but some, including many cloud-based backup systems, do not. Considering that recovering deleted or overwritten files is one of the most common reasons for restoring data from backup, your backup software should have this important feature.

Open File Backup

In most circumstances, backing up open files (files currently in use) is risky, at best. The backup may fail, or the files may not be in a useable state when they are restored. This is especially true of database files, where multiple data files and index files need to be kept in synchronization. Many database systems provide utilities or *application programming interfaces* (APIs) to put the files in a safe state for backup. Utilities are loaded before the backup, while APIs are used directly by the backup software. These utilities and APIs lock the database to prevent updating during backup. Any writes to a protected file during backup are written to a buffer, or temporary data location. Read requests can access the buffer, so you'll be able to get to any data written during backup, even while the database is locked. The database is updated from the buffer after backup.

On Windows systems, many databases use Microsoft's *Volume Shadow Copy Service* (VSS) for this purpose. VSS is a set of services that are designed to provide consistent copies of Windows systems and Windows applications, such as Microsoft SQL Server and Exchange. The VSS API is available for other applications to use as well.

The most common database used on Macs is Filemaker Pro. The server version provides facilities through the admin console to back up open databases, but the desktop version does not; the program should always be closed before backing up. If you want to include Filemaker Pro Server in your regular backups, only include the backup files created through the admin console, not the live, running database.

Many databases running on Linux provide scripts or utilities that allow the safe backup of open files, usually as part of a regular backup routine. As with other operating systems, backing up open files without these mechanisms will either fail or potentially produce corrupted backups.

Compression and Deduplication

Compression and deduplication are two techniques for reducing media space requirements. Compression is commonly used on primary storage as well as backup and archiving media. Deduplication is primarily used on backup and archiving media. Compression operates on the bit level and removes redundant bits of data. It replaces them with codes that can be used to restore those bits when the data is read. Deduplication is similar to compression in that it reduces the amount of space a given data set requires on disk.

A deduplication program analyzes data and looks for files or blocks of data (depending on the particular deduplication method employed) that are the same. When two or more files or blocks match, the system sets a pointer to a single file or block and does not store multiple copies of that data. Deduplication provides the greatest benefit where there is a significant amount of redundant data. On primary media, compression can be implemented on a volume basis or by individual file and folder. With backups, compression is typically applied to the entire backup job. This can be done through backup software or, in the case of magnetic tape drives, implemented in firmware. Deduplication is often included as a feature of enterprise-level backup systems as well as some systems designed for smaller organizations. Deduplication can also be used on primary storage systems, and can be implemented at the file system level. The SDFS file system for Linux from OpenDedup (opendedup.org) has some deduplication capability built in. Deduplication is also planned for a future release of the BTRFS file system.

Both compression and deduplication can save disk space, but without hardware dedicated to their processing, they will almost always negatively impact performance. They are of limited advantage for most photographers because much of our image data does not benefit from either space reduction technique.

Encryption

Encryption is employed to prevent unauthorized entities from viewing and copying your data. Encryption uses a *key*, which is a piece of information that is used to translate your data into unreadable gibberish and back again. In simplest terms, the length and complexity of the key determines the difficulty of breaking the encryption. Longer keys also have greater impact on performance. 56-bit keys used to be common, but 128 bits is the new usual minimum for data encryption, with 256-bit (and larger) keys becoming more common. Keys are generated with a random key generator.

No encryption scheme is completely unbreakable, but the time and cost of breaking modern encryption algorithms, especially 256-bit keys or higher, are enough to deter all but the largest organizations with deep pockets (think governments) from even attempting decryption.

There are three primary methods for encrypting backups:

- **Software-based encryption:** Many backup software packages include encryption capability. Employing software-based encryption will usually affect backup and restore performance.
- **Encryption appliances:** Encryption appliances are typically expensive and used with enterprise-class backup systems. Because the appliance has its own processor, encryption and decryption usually have no negative impact on backup and restore performance.
- **Tape drive encryption:** LTO-4 and later generation LTO tape drives include hardware-based encryption. Tape drive encryption does not impact performance. Encryption and the encryption key are controlled through the tape drive's interface and managed by the backup software. If the drive is part of a tape library, encryption is managed by the library's management system.

Who Maintains the Keys?

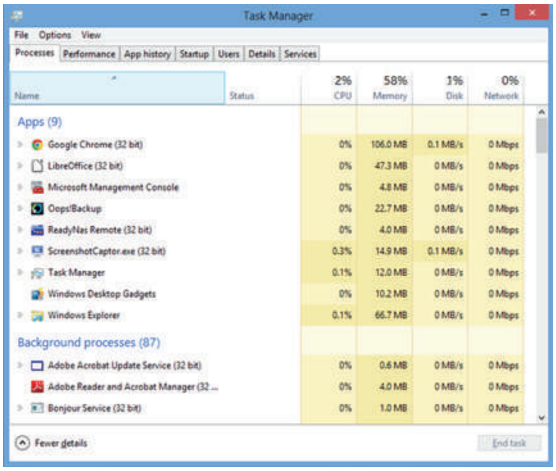
Sometimes the user maintains the keys, and sometimes the application handles it all internally. In this last case, the user typically supplies a login name and password. Remember: if you lose the key or password, you lose access to your data. This is the danger of using backup encryption.

Requirements for an Effective Backup and Archiving System

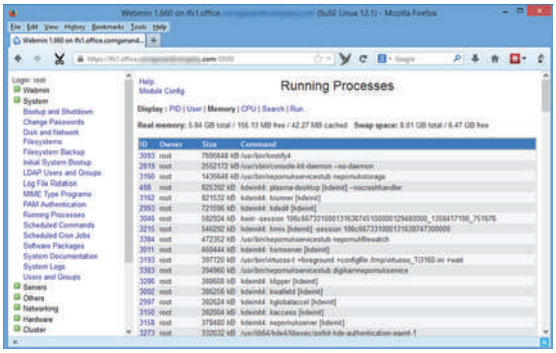
There is no one best approach to backing up and archiving data because everyone has different requirements. You may need more than one type of backup software or system to meet all your backup and archiving goals. Here are some things to look for in your backup system:

- **The ability to create clone or image backups for fast system recovery.** Depending on your operating system and the backup software you are using, there are numerous ways to accomplish this. There are several programs for Macs that will maintain a bootable clone on a continuous or near-continuous basis. For Windows and Linux computers, there is software that can be used to periodically create image backups, including incremental image backups and clones. Your imaging or clone software should allow you to back up a single partition and multiple partitions.

- **The ability to select by volume, file, or directory.** Your file backup software should allow you to select what to back up by the criteria most important to you.
- **The ability to create multiple backup jobs and back up to multiple destinations.** You might want to back up some data to a local destination and other data to the cloud, or you might want the same backup written to two different locations.


Task Manager
File Options View
Processes Performance App history Startup Users Details Services
Fewer details
End task

Name	Status	2% CPU	58% Memory	1% Disk	0% Network
Apps (9)					
Google Chrome (32 bit)		0%	106.0 MB	0.1 MB/s	0 Mbps
LibreOffice (32 bit)		0%	47.3 MB	0 MB/s	0 Mbps
Microsoft Management Console		0%	4.8 MB	0 MB/s	0 Mbps
Oops!Backup		0%	22.7 MB	0 MB/s	0 Mbps
ReadyNas Remote (32 bit)		0%	4.0 MB	0 MB/s	0 Mbps
ScreenshotCaptor.exe (32 bit)		0.3%	14.9 MB	0.1 MB/s	0 Mbps
Task Manager		0.1%	12.0 MB	0 MB/s	0 Mbps
Windows Desktop Gadgets		0%	10.2 MB	0 MB/s	0 Mbps
Windows Explorer		0.1%	66.7 MB	0 MB/s	0 Mbps
Background processes (87)					
Adobe Acrobat Update Service (32 bit)		0%	0.6 MB	0 MB/s	0 Mbps
Adobe Reader and Acrobat Manager (32 ...		0%	4.0 MB	0 MB/s	0 Mbps
Bonjour Service (32 bit)		0%	1.0 MB	0 MB/s	0 Mbps



Running Processes

Display : PID | User | Memory | CPU | Search | Run..

Real memory: 5.84 GB total / 156.13 MB free / 42.27 MB cached **Swap space:** 8.01 GB total / 6.47 GB free

```
ID    Owner  Size        Command
3093  root   7695648 kB  /usr/bin/knotify4
2819  root   2552172 kB  /usr/sbin/console-kit-daemon --no-daemon
3160  root   1435648 kB  /usr/bin/nepomukservicestub nepomukstorage
488   root   825392 kB   kdeinit4: plasma-desktop [kdeinit] --nocrashhandler
3162  root   821532 kB   kdeinit4: krunner [kdeinit]
2993  root   721596 kB   kdeinit4: kded4 [kdeinit]
3046  root   582924 kB   kwin -session 106c6673310001316387451000001294800000_1358417190_751676
3215  root   546292 kB   kdeinit4: kmix [kdeinit] -session 106c6673310001316387473000000
3384  root   472352 kB   /usr/bin/nepomukservicestub nepomukfilewatch
3011  root   450444 kB   kdeinit4: kamserver [kdeinit]
3193  root   397720 kB   /usr/bin/virtuoso-t +foreground +configfile /tmp/virtuoso_TJ3160.ini +wait
3383  root   394960 kB   /usr/bin/nepomukservicestub digikamnepomukservice
3206  root   388668 kB   kdeinit4: klipper [kdeinit]
3002  root   386256 kB   kdeinit4: kwalletd [kdeinit]
2997  root   382624 kB   kdeinit4: kglobalaccel [kdeinit]
3150  root   382004 kB   kdeinit4: kaccess [kdeinit]
3158  root   378480 kB   kdeinit4: nepomukserver [kdeinit]
3273  root   332032 kB   /usr/lib64/kde4/libexec/polkit-kde-authentication-agent-1
```

- **The ability to span multiple backup targets.** This is particularly important if you are using removable media, such as tape or removable disks, for backup. If one backup target gets full, you should be able to continue the backup by replacing that target with new media.
- **Simple and straightforward restoration.** Backup is done calmly and restoration is done in a panic. Complex restoration procedures do not help.
- **The ability to maintain backups of deleted files and multiple versions of changed files.** One of the most common reasons for restoring data is that a file has been deleted or overwritten by a newer version.
- **For backup software that maintains multiple versions, the use of delta encoding to only record file changes instead of making multiple copies of entire files is a plus.** This not only saves space on backup media, it also improves backup performance.

Backup Tips

- **Create a recovery disk.** Depending on your operating system and backup software, this could be a CD, DVD, or USB flash drive.
- **If you are imaging your system disk, consider creating a separate data partition, or store your data on a different disk.** Use a smaller partition (or partitions) for the operating system and application programs, and a separate partition (or disk) for data. This allows you to image the system partitions and back up the data partition separately. This makes backup and restoration of the system from an image file faster and easier. It also makes it less likely that restoring the image will overwrite newer data.

- **Temporary files and browser cache files should be excluded from your backup.** See Common Browser Default Cache File Locations. On Windows, you should exclude **C:\Users\<username>\AppData\Local\Temp**. You can exclude Lightroom previews since they are easily recreated. If your backup program maintains multiple file versions, you may want to consider excluding Lightroom catalog backups, as well.

> *Lightroom catalog backups refer to the backups created by Lightroom, not by your backup software. Lightroom backs up my catalogs, but I exclude those files from my regular file and folder backups, since my backup software maintains multiple versions of the catalog anyway.*

- **Label all removable media**. Backup media should be labeled in a manner that is understandable to anyone who might be involved in backing up or restoring data. The label information should include basic information about what has been backed up, the type of backup (full, continuous, incremental, etc.), and the date of the last backup (perhaps written in pencil). If you use a media rotation scheme, the media's position in the rotation should be included, as well.
- **Make sure you include program templates, profiles, and configuration files with your file backups.** See Common Default Template and Configuration File Locations.

Common Default Browser Cache File Locations

Safari (Mac)

Cache files: ~/Library/Caches/com.apple.safari

Internet Explorer 8, 9, and 10 (Windows 7 and 8)

C:\Users\<username>\AppData\Local\Microsoft\Windows\Temporary Internet Files

Firefox (Mac)

~/Library/Caches/Firefox/Profiles/<profile>/Cache

Firefox (Windows 7 and 8)

C:\Users\<username>\AppData\Local\Mozilla\Firefox\Profiles\<profile>\Cache
C:\Users\<username>\AppData\Local\Mozilla\Firefox\Profiles\<profile>\OfflineCache

Firefox (Linux)

~/.mozilla/firefox/<profile>/Cache

Chrome (Mac)

/Users/<username>/Caches/Google/Chrome/Default/Cache/

Chrome (Windows 7 and 8)

C:\Users\<username>\AppData\Local\Google\Chrome\User Data\Default\Cache

Chrome (Linux)

/home/<username>/.config/google-chrome/Default/Application Cache/Cache

Common Template, Profile, and Configuration File Locations

Please note that some of this data may be in hidden folders or files. See your operating system documentation for information on how to access hidden files and folders.

Aperture

~/Library/Application Support/Aperture

Lightroom 4 and 5 (Mac)

Preferences: //Users/<user name>/Library/Preferences

Presets and Templates: //Users/[user name]/Library/Application Support/Adobe/Lightroom

Registration Data:
4.0: //Library/Application Support/Adobe/Lightroom/Lightroom 4.0 Registration
5.0: //Library/Application Support/Adobe/Lightroom/Lightroom 5.0 Registration

Lightroom 4 and 5 (Windows 7 and 8)

Preferences: C:\Users\[user name]\AppData\Roaming\Adobe\Lightroom\Preferences

Presets and Templates: C:\Users\[user name]\AppData\Roaming\Adobe\Lightroom

Registration Data:
4.0: C:\ProgramData\Adobe\Lightroom\Lightroom 4.0 Registration.lrreg
5.0: C:\ProgramData\Adobe\Lightroom\Lightroom 5.0 Registration.lrreg

Photoshop CS5 and CS6 (Mac)

Files may be in subfolders.
General settings, actions, Camera Raw preferences, third-party plug-in settings, editing and painting tools, paths, save for web, filters and effects, lens profiles, workspaces: Users/<username>/Library/Preferences/

Custom color and proof settings, saved presets: Users/<username>/Library/Application Support/Adobe

Photoshop CS5 and CS6 (Windows 7 and 8)

Windows registry settings may be in sub-keys.
General settings, actions, custom proof setups, custom color settings, editing and painting tools, save for web, filters and effects, lens profiles workspaces, saved presets: Users/<username>/AppData/Roaming/Adobe

Camera Raw registry settings: HKEY_CURRENT_USER/Software/Adobe/CameraRaw/6.0
Users/[user name]/AppData/Adobe/CameraRaw/Settings

Paths registry settings, Third-party plug-in registry settings: HKEY_CURRENT_USER/Software/Adobe/Photoshop

Corel AfterShot Pro (Mac)

Settings: ~/Library/Application Support/AfterShot Pro

Catalog: ~/Pictures/AfterShot Pro Catalog

Corel AfterShot Pro (Windows)

Settings: C:\Users\<username>\AppData\Local\Corel\AfterShot Pro
C:\Users\<username>\My Pictures\AfterShot Pro Catalogs

Corel AfterShot Pro (Linux)

Settings: ~/.AfterShot Pro

Catalog: ~/Pictures/AfterShot Pro Catalogs

Internet Explorer

Favorites: C:\Users\<username>\Favorites

Firefox (Mac)

~/Library/Mozilla/Firefox/Profiles/<profile folder> ;
~/Library/Application Support/Firefox/Profiles/<profile folder>

Firefox (Windows 7 and 8)

C:\Users\<username>\AppData\Roaming\Mozilla\Firefox\Profiles\<profile folder>

Firefox (Linux)

~/.mozilla/firefox/<profile folder>

Chrome (Mac)

/Users/<username>/Library/Application Support/Google/Chrome/Default/Preferences

Chrome (Windows 7 and 8)

C:\Users\<username>\AppData\Local\Google\Chrome\User Data\Default\Preferences

Chrome (Linux)

~/.config/google-chrome/Default/Preferences

Backing Up NAS and SAN Configurations

Most NAS and SAN devices allow you to back up your array configuration settings to a file so you can restore them in an emergency. It's a good idea to back up your settings both before and after any configuration changes. When you save the file, include the date and time in the file name, and save it to a location on a device other than the SAN or NAS. Make sure those files are included in your normal backup routine.

Backing Up Your Website

If you have a website hosted by an Internet service provider (ISP), you should be backing it up. Don't rely on the ISP to do this for you. Always make sure you keep local copies of anything you post on the web, and back up those copies as part of your regular backup routine. Even though you have this data backed up, you still want to back up your complete website so that you can quickly recover it in an emergency or, if necessary, move to a different hosting company.

There are two types of hosting you may be using. The first type is a generic site, where you (or your web designer) select the web-building tools (WordPress, Joomla, etc.) and build the site. The second type is a "we-do-it-all-for-you" type of site. Most photo hosting sites fall into this category. The hosting company either builds your site with their site-building tools, or they provide a limited range of layout options, and you add your images and data.

If you have a generic-type site, you can usually back it up using *file transfer protocol* (FTP). A command-line version of FTP ships with pretty much every operating system. There are also GUI FTP clients, such as WinSCP for Windows (winscp.net) or Cyberduck for Mac OS and Windows (cyberduck.ch), and there are browser plugins like FireFTP for Firefox (fireftp.net). Your web hosting service may supply tools to help you back up your web files. If you are not a seasoned web developer, you may need technical assistance to help figure out exactly what to back up.

One advantage of the command-line FTP versions is that they are usually easy to script and schedule for automatic backup. Typically you will back up to your local computer and include those files in your normal system backups.

If you use a "we-do-it-all-for-you" web-hosting service or a photo-hosting service, ask (preferably before you sign up) how to back up your web data. Also find out what would be involved in moving to another service, which can be an issue with this type of web service. If the web development tools are proprietary, you may be locked in. This applies to photo sites as well. You may be able to retrieve your photos, articles, blog posts, etc. (you should already have local copies of all this anyway), but you may not be able to back up or retrieve the organizational metadata that defines your site layout. If you want to move to a different service, you may end up rebuilding everything from scratch. Find out this information before you invest a lot of time and money into your website.

Archiving Data

Archiving is the process of moving data that is no longer used to a separate storage device for long-term retention. Of course, there is no reason that you cannot archive copies of currently used data as well. When archiving data you need to consider what to archive, the retention period, and maintaining archive integrity.

What to Archive

Not all your data needs archiving. For example, software, system image files, and system configuration information is typically of short-term value. It needs to be backed up, but not archived. Images, financial records, and other business records usually do need to be archived. In many industries, certain data, such as email, needs to be archived due to regulatory requirements. Even without regulatory requirements, archiving email may be valuable (or possibly detrimental) in the event of litigation. You (and your business advisers, if you have them) need to make the determination as to what needs to be archived.

Retention Period

For most photographers, the retention period for their photographic images is open-ended. Images may end up having artistic or historical value, and with current U.S. copyright laws, they may have monetary value to the photographer's heirs.

Some data, such as tax records, might be archived for a specific period of time. When archiving such data, you should have a mechanism for destroying it at the end of the archive period. This means separate archive media for data with different destruction dates. Also, any media with a destruction date should be labeled with that date.

Maintaining Archive Integrity

The first thing to consider is the expected life of the archive media. First, quoted lifespan figures for various types of media are, at best, statistical averages based on previous experience, accelerated age testing, and other factors. At worst, they are little more than marketing spin. In any case, there is always the chance that any specific unit of storage media can fail prematurely. You should also consider how long the hardware and software required to read that media will be available. For example, even if I still had a hard disk from a thirty-year-old IBM PC-XT, I would have a hard time finding an ST-506 disk interface to connect it to, much less a working PC-XT.

Another concern is the future readability of file formats. This should be a major concern for proprietary RAW images and old word processing formats, for example. Archived media should be checked every few years for readability. Do a full restore to disk (not to the original location), if possible, and at least spot-check the files. Archives should be moved to new media at least every five to eight years. You should also make sure software is available to read your archived file formats.

Selecting Media for Archiving

Keep in mind that the purpose of archiving is long-term storage, while the purpose of backup is restoring lost data. NAS and SAN devices, as well as running hard disks, may be appropriate for first-line backup, but they do not usually make the most sense for long-term archiving because they are subject to power-related problems, viruses and malware, and accidental file deletion or modification. However, they do have the benefit of allowing you immediate access to your archived data.

Online storage has similar issues. The difference is that someone else is in charge of the day-to-day maintenance of the systems that store your data, but the data is still stored on spinning hard disks, and you are effectively paying to keep those disks spinning. You may not be in charge of the day-to-day maintenance, but you are still responsible for the data. If you read the service level agreement with your online service provider, you will likely find no guarantees—just a promise to use their best efforts to keep your data safe. Even the biggest cloud service providers have lost user data. Others have ceased operation, either voluntarily or otherwise, taking user data with them. The best option for archiving data is some type of removable media, which includes optical discs (CD, DVD, Blu-ray), hard disks (external or removable), flash media, or digital magnetic tape.

Optical Disc

Experience has shown that data on writable CDs and DVDs can become unreadable in as little as two or three years. Because of the relative newness of Blu-ray, the jury is still out on its long-term ability to retain data. Capacity and write speed are also issues, as is media cost per GB, especially with Blu-ray. Readers for optical media are common, and many computers ship with drives that can read all three formats, so access to data is quick and easy. However, few readers currently shipping with consumer-grade computers can read 100 GB BDXL disks.

External and Removable Hard Disks

Hard disk manufacturers do not specify a shelf life for data stored on non-operating hard disks, but the expected life of a continuously running hard disk averages around five years. There is some evidence that data on non-running disks can degrade in two or three years. One good thing about hard disks for archiving is that you don’t need special hardware, such as a tape drive, to retrieve your data.

Flash Media

Flash media, including SSDs and USB flash drives, appears to be very stable. When stored properly, it should be able to retain data for ten years or more. The biggest issue with flash media is the cost.

Digital Magnetic Tape

Magnetic tape, in particular LTO tape, has demonstrated its ability to store data for extended periods of time. The biggest downside is the cost of the tape drive needed to write the data, and the fact that if you need to recover data years from now, you’ll need a compatible tape drive to read that tape. Fortunately, LTO tape drives read three generations of tapes, so an LTO-6 drive reads LTO-4, LTO-5, and LTO-6 tapes. LTO-3 tape drives, which are still currently available, will read LTO-1 tapes from over twelve years ago. One more advantage of tape—it is immune to malware and viruses. The data stored on tape may be infected, but unlike hard disks, the tape itself cannot become infected.

Archive File Formats

Because you are archiving for long-term storage, archived data should be stored in a manner that does not require proprietary software to read it, since such software might not be available when you need to extract the data. It's best to store archived data in its native format. If you are archiving to LTO tape, use the LTFS file system, which will allow the data to be read by any compatible drive with LTFS support.

What about compression? Compression does not significantly reduce the size of photographic image files. If you have software or tape drives that compress by default, I suggest you turn compression off. If you choose to encrypt your archives, make sure that the passwords or encryption keys will be available years down the road. Otherwise you're putting an unbreakable lock on something that may have artistic, historical, or personal value.

Storing Archive Media

All digital media should be stored in a clean environment with moderate, stable temperature and humidity. All media should be kept in protective cases. Tapes and removable disk cartridges usually come with plastic cases, and cases for bare hard disks are available online and in electronics stores. Maintaining consistent, moderate humidity is the most important factor for preserving the life of hard disks and tapes. Optical media should be protected from scratching and light. Dye is used to differentiate between logical ones and zeros (the laser actually burns the dye layer), and light will fade the dye over time.

Protective disk storage boxes are available from local electronics stores and online sources

At least two copies of archived media should be maintained in at least two physical locations, preferably far enough apart that a natural disaster would be unlikely to affect both sites.

Backup Software

There are hundreds of backup software packages available, and this is not a comprehensive review of all of them. I am profiling software I've had personal experience with, or that has been highly recommended to me by people I trust. I've profiled products that I believe are suitable for the single user or small office, although some may be suitable for larger organizations, as well. These products are all reasonably priced—typically between $25-80 per desktop.

The one open source product profiled, Mondo Rescue, is available at no cost, and Apple Time Machine and the Microsoft-supplied backup software are bundled with the operating systems.

Macintosh

Apple Time Machine

Time Machine, which is included with current versions of Mac OS X, creates incremental backups of all locally attached disks by default. Specific disks, volumes, folders, or files can be excluded. Current versions of Time Machine can back up data to multiple types of targets, including external disks and NAS devices that include Time Machine support. Time Machine automatically makes hourly backups that it saves for 24 hours. It saves daily backups for a month, and weekly backups older than a month until the backup target runs out of space. Time machine can restore files, folders, and entire systems. Time Machine allows you to browse backups to restore previous versions. Time Machine does not record delta changes—it backs up complete changed files. Time Machine doesn't directly support off-site backups, but at least one third-party NAS device provides off-site capability for Time Machine. Other than file, folder, and volume exclusions, Time Machine is not very configurable. For example, the scheduling is pretty much fixed, and the number of backups saved cannot be changed. Time machine is not suited for creating long-term archives.

Apple Disk Utility

Disk Utility is Mac's general-purpose disk-management utility. It can be used to repair permissions, erase disks, and set up RAID if you have multiple disks. It can also copy the contents of a hard disk to another location (this function is called *Restore*) and create disk images. The Restore function can be used to clone hard disks, including bootable system disks, while resizing partitions in the process, allowing for migration to a larger disk.

Support

Apple provides 90 days of software telephone support. Additional no-cost support may be available depending on the specific warranty coverage for your product.

SuperDuper

SuperDuper (shirt-pocket.com) is primarily a disk-cloning program. It allows you to create a bootable clone of your system. There is a free version, with limited capabilities, and a paid version, profiled here. SuperDuper is an online cloning program, which means you don't have to shut your system down to create the clone. Its *Smart Update* feature allows for updating existing backups manually or on a scheduled basis. Smart Update only makes the changes, additions, and deletions necessary to make the backup match the source. SuperDuper will also let you create image files, which they call Sparse Images, side-by-side with other data on external or network drives. Sparse Images are not bootable, but can be restored to a drive in bootable condition. SuperDuper lets you create scripts to exclude files from backup. This allows the exclusion, for example, of temporary files that do not need to be part of a system restore. SuperDuper does not maintain previous file versions or deleted files.

Support

SuperDuper provides unlimited email support.

Carbon Copy Cloner

Carbon Copy Cloner (CCC) (bombich.com) creates bootable clones of Macintosh disks and will also back up files and folders to external disks and other Macs on your network. CCC has a function that allows you to update your backups on a scheduled basis. CCC incremental backup offers the option of maintaining deleted files or previous versions of changed files in a separate archive folder on the backup target. It also allows you to control the size of the archive by *pruning*, or deleting older files based on available disk space, the size of the archive, or the age of the files. The CCC archive function maintains complete files, not delta changes. The CCC archive should be considered temporary storage only if automatic pruning is used. The user manual specifically recommends maintaining a copy of your archive folder elsewhere.

Support

CCC provides online help desk support.

ChronoSync

ChronoSync (econtechnologies.com) is a cloning and synchronization program. It can create bootable clones of disks and keep them synchronized; it can synchronize sets of folders and files; and it can perform bidirectional synchronization, keeping two active sets of files in synchronization. After an initial synchronization, ChronoSync monitors file changes on the source system and replicates only those changes to the target system on a scheduled basis. ChronoSync also provides versioning (which their documentation calls *archiving*), which allows you to maintain deleted files and older versions of changed files. The user can specify a minimum and maximum number of versions to keep, as well as a maximum version age. The ChronoSync software lets you browse, open, or restore versions. ChronoSync's versioning keeps complete files, not delta changes, so it can require a considerable amount of disk space over time.

Support

ChronoSync includes unlimited email support and free upgrades.

CrashPlan

CrashPlan (crashplan.com) is an online backup service. It provides free backup software that can be used with or without online backup. CrashPlan's software lets you created multiple backup jobs and back up data to multiple locations, including local disks, network shares, remote computers and Crashplan's online backup site, CrashPlan Central.

CrashPlan includes versioning, compression, and deduplication by default, all of which are configurable. It will back up Macs, Windows, and Linux. If you have a CrashPlan online account, you can access data that is backed up to CrashPlan Central from a mobile device. Backup is based on schedules, and scheduling is completely configurable. You can also set bandwidth limits for local and Internet traffic. CrashPlan has multiple plans for home and business, and allows customers (U.S. only) to seed the cloud account by sending data on a hard disk. Crashplan's software always encrypts backed up data, even when you back up to a local device. One caveat with CrashPlan—if you remove a backup data source, the data previously backed up from that source will be removed as well.

Support

CrashPlan provides web, chat, email, and phone support to paid users.

Retrospect

Retrospect was once the leading backup software for Macs, but due to multiple corporate acquisitions (Dantz, the company that produced Retrospect, was first acquired by EMC, then sold to Roxio, which was acquired by Rovi), the product languished and fell behind its competitors. In 2011 a group of the original developers bought the product from Rovi and formed Retrospect, Inc. (retrospect.com) to continue developing the product.

Retrospect is a very powerful program. It is sold in a desktop edition and multiple server editions. Regardless of the edition, the program has two parts—the engine and the client. Retrospect can back up multiple clients to the system hosting the engine. Retrospect for Mac (as well as for Windows) can back up Mac, Windows, and Linux clients. The desktop version includes five client licenses (the computer hosting the backup engine and four additional clients). Retrospect will perform traditional file and folder backups, and can create images and bootable clones. In file and folder mode, Retrospect provides file versioning with extensive control, although it currently does not record delta changes—it keeps multiple versions of complete files. Retrospect for Mac supports tape drives in native mode, but does not currently support LTFS. It lets you create and schedule multiple backup jobs and provides extensive support and documentation for disaster recovery. Its Duplicate mode is great for archiving data, because it copies data in native format with a wide range of selection criteria. However, its terminology and wide range of file selection options make the user interface a bit complex, so it takes some getting used to.

Support

Retrospect provides free phone and email support during a 45-day trial period, and 30-day phone and email support after purchase. Online forums, knowledgebases, and tutorials are always available for free. Annual support contracts are also available.

Windows

Windows 7 Backup and Restore

Windows 7 Backup and Restore allows you to create a recovery disk (CD, DVD, USB flash drive, etc.), create system images, and perform file backups. Also, you can restore files from the system image backup. It supports a single backup schedule for both file and image backups. If you want to change the list of files to be backed up or the backup times, you have to overwrite the existing schedule. New backups overwrite existing backups (unless you change the backup media), so versioning is not supported. You can designate the disk or network share for the backup, but you cannot designate a particular folder. Microsoft does not support restoring images created by Windows 7 Backup to other computers.

Support

Windows support policies apply.

Windows 8 File History (with Windows 7 File Recovery)

File History is a scheduled backup program that backs up files stored in Libraries, Desktop, Favorites, and Contacts folders. The only way to back up other folders is to create a new library or add the folders to an existing library. You can also exclude folders. The default schedule is every hour, but that can be changed. File History maintains multiple versions of files, and the retention period can be changed. The default is *forever*, but other options include one, three, six, and nine months, one or two years, or until space is needed. You cannot set the number of versions to retain. File History can use external drives or network shares as targets, but will not write to CDs, DVDs, or Blu-ray discs. Windows 8 also includes the same functions as Windows 7 Backup and Restore, but it has been renamed Windows 7 File Recovery and is accessible from a link at the bottom of the File History page.

Support

Windows support policies apply.

Windows 8.1 File History (with System Image Backup)

File History in Windows 8.1 differs from the 8.0 version in that it will no longer create a System Repair CD, only a System Recovery USB drive. Windows 7 File Recovery is also gone; the link has been replaced with a link to System Image Backup. Unfortunately, System Image Backup has no scheduling facility, but a web search for “how to schedule a Windows 8.1 System Image backup” will return several results that show how to use the Windows Task Scheduler.

Support

Windows support policies apply.

Altaro Oops! Backup

Altaro Oops! Backup (altaro.com) is an easy-to-use file and folder backup program that provides automatic versioning. It detects file changes and then backs up the changed files on a schedule set by the user. Oops! Backup uses what Altaro calls *ReverseDelta Technology*, meaning it only backs up actual changes to files. The latest backed up version of a file is stored as the complete file, while deltas are stored for previous versions. This allows you to retrieve the latest version quickly, and since it is not stored in a proprietary format, you can retrieve it using standard copy utilities, if necessary. Oops! Backup lets you keep a full copy of each file after a specified number of versions are saved. You can also set a time limit after which old versions will be purged. Oops! Backup will back up to and from external disks, NAS devices, and network shares. For laptops (or desktops) Oops! will automatically back up when the backup drive is connected. You can also set up a second backup drive or location that will synchronize daily with your primary backup, at a time the user specifies. Oops! Backup is easy to set up and use. Because of its ReverseDelta design, it cannot back up to tape, but a backup disk can be copied to tape for archiving purposes. A limitation of Oops! Backup is that it can only back up a single backup job on a single schedule.

Support

Altaro offers email support, live chat, user forum, and a knowledgebase.

Altaro BackupFS

Altaro BackupFS (altaro.com) is similar to Oops! Backup, but has a number of additional features. First, it backs up Windows servers, while Oops! only runs on Windows workstations. Altaro runs as a Windows service so it can run without a user logged into it. It provides greater scheduling flexibility and it sends email notification on backup success or failure.

Support

Altaro offers email support, live chat, user forum, and a knowledgebase.

CrashPlan

See CrashPlan, under Macintosh.

Macrium Reflect

Macrium Reflect (macrium.com) is a versatile backup program that provides disk imaging, cloning, and file and folder backup. Image files are compressed and can be mounted as drive letters. Full, incremental, and differential images can be created. A system recovery disk based on WindowsPE, Windows Automated Installation Kit, or Linux can be created to recover a system disk image. Reflect’s file and folder backup creates a compressed virtual drive (in .zip format) that can be mounted as a drive letter. Additional full, incremental, and differential backups each create and store their data in new .zip files. A system recovery disk based on WindowsPE, Windows Automated Installation Kit, or Linux can be created to recover a system disk image. The Professional and Server editions allow image redeployment to new computer hardware. A free edition that does imaging and cloning, with scheduling, is also available. Unfortunately, Reflect's compression will not significantly reduce the disk space required by most photographic image files. If you want to store backups on LTO tape, the backup should first be written to disk, then copied to tape.

Support

Support is via email, user forums, and a knowledgebase. The free version is not supported.

Retrospect

See Retrospect, under Macintosh. The Mac and Windows versions are very similar, but in addition to supporting native tape mode, the Windows version can duplicate (copy) to LTFS tapes, although LTFS doesn't support extended attributes of the Windows NTFS file system. This is a good option for archiving data to tape in native mode.

Linux

Most backup programs for Linux are free and open source, but they require knowledge of the Linux command line. The programs profiled here have a menu or GUI interface for most common operations.

Mondo Rescue

Mondo Rescue (mondorescue.org) is an online imaging program for Linux. It will back up to local disk, external disk, NAS and SAN devices, network locations, and tape. It creates a bootable recovery disk and allows complete bare-metal restoration. It also allows the restoration of individual volumes, partitions, files, and directories. Mondo is free and open source software and is used in small offices, government agencies, and large corporations. Mondo is primarily a command-line program, but it also has a basic menu interface.

Support

Support is available from the Mondo website, wiki, and mailing list archives. Paid commercial support is available as well. Information about support options is available here: mondorescue.org/support.shtml.

Areca Backup

Areca Backup (areca-backup.org) is a Java-based backup program for Linux and Windows. It is a free and open source program that performs full, incremental, differential, and delta backups. It is relatively easy to install and use once you understand the terminology and interface. (I highly recommend following the tutorial at areca-backup.org/tutorial.php to get started.) Files are stored in original file-and-folder format or in a standard .zip file. Scheduling requires using the Areca command-line interface to create backup scripts that can be scheduled using the cron scheduling utility.

Support

Support is limited to the forums on the Areca website.

Rsync and rsync-based Backup Applications

Rsync (rsync.samba.org) isn't technically a backup program—it's a file synchronization program. On its own, it's a command-line utility that can be scripted to replicate files from one computer system to another. When it updates existing files it only records delta changes, minimizing synchronization time and network traffic. Rsync is available for all commonly used operating systems, so it's not just a Linux tool. Rsync is also the basis of a number of open source backup and synchronization tools; there are GUI front-end programs, such as Grsync (opbyte.it/grsync/) for Linux, Mac, and Windows. Using rSync, even with a GUI front end, can be a complex process. It's not for those who are unfamiliar with command-line programs.

A number of backup applications are either based on rsync or they use rsync functions. These include Back In Time (backintime.le-web.org), TimeVault (launchpad.net/timevault), and FlyBack (flyback-project.org). Although these are all GUI-based utilities, they all have their own quirks and may, at some point, require command-line-based Linux utilities. All are free and open source (although the authors may ask for donations) and all are a little rough around the edges.

Support

Rsync support is pretty much a do-it-yourself affair. The web page will point you to documentation and the rSync mailing lists. Support for rsync-based backup programs is similar. See their websites for details.

Retrospect

The Retrospect backup engine only runs on Mac and Windows computers, but it does include client software for Linux that allows the Linux system to be backed up across the network. Retrospect Desktop Edition includes licenses for backing up five computers. If you also use Mac OS or Windows computers, Retrospect makes it easy to add your Linux system to your backup routine. Unlike most other Linux backup programs, using Retrospect doesn't require knowledge of the Linux command line.

A Basic Mac Backup Routine

For a single Mac, here is a basic routine:

- Use Time Machine to back up data to an external drive.
- Use cloning software (ChronoSync, SuperDuper, Carbon Copy Cloner, etc.) to maintain a bootable clone.
- Back up critical data to a remote or cloud location, preferably with file versioning. This can be done a number of ways, including:
    - cloud services and software like CrashPlan (Crashplan actually lets you back up to a friend's computer at a remote location)
    - private cloud, as provided by NAS vendors like LenovoEMC and NETGEAR
    - Rsysnc-based solutions (this is trickier, because it typically requires knowledge of routers and firewalls); Rsync ships as part of Mac OS, although newer versions are usually available from rsync.samba.org

A Basic Windows Backup Routine

For a single Windows PC, here is a basic routine:

- Use third-party software to back up data, preferably with versioning, on a continuous, near-continuous, or scheduled basis.
- Use the imaging function of the included Windows backup software or third-party software to periodically create a restorable image of your system. It's a good idea to create a separate disk partition for data or store your data on a separate disk. That way you can exclude the data partition from the image backup, creating a much smaller file.
- Back up critical data to a remote or cloud location, preferably with file versioning (see *A Basic Mac Backup Routine*).

A Basic Linux Backup Routine

For a single Linux system, here is a basic routine:

- Use a Linux file-based backup program such as Areca to back up data on a scheduled basis.
- Use Mondo Rescue to periodically image the partitions that are required for booting the system.
- Back up critical data to a remote or cloud location, preferably with file versioning. This can be done a number of ways, including:
    - cloud services and software like CrashPlan (Crashplan actually lets you back up to a friend's computer at a remote location)
    - private cloud, as provided by NAS vendor LenovoEMC (NETGEAR does not provide a Linux version of ReadyNAS Remote)
    - Rsysnc-based solutions (this is trickier, because it typically requires knowledge of routers and firewalls)

My Backup Routine

My office is in my home, and both my wife and I work there. My backup routine may seem like overkill, and it probably is. However, I have not found a single backup system that does everything I want. Our complement of systems consists of:

- my desktop Windows PC with external hard disk for photographic images;
- my Windows laptop;
- my wife's desktop Windows PC;
- a Linux server with attached SAN;
- a remote NAS for off-site backup (private cloud); and
- an LTO-6 tape drive.

I use five backup programs:

- Windows Backup
- Mondo Rescue, free and open source imaging software for Linux
- Altaro Oops! Backup
- CrashPlan
- Retrospect

As much as possible, my backups are automated. The bundled Windows backup software is used to image the Windows PCs, and Mondo Rescue is used to image the Linux server.

I use Crashplan to back up data from all systems on an automated basis. Except for photographic images, all data is backed up to the CrashPlan cloud service. All data, including photos, is backed up to the server-attached SAN.

Altaro Oops! Backup backs up critical data and photos from my PC, my wife's PC, and the Linux server, and saves the data to an external hard disk attached to my PC. Oops! then synchronizes the backup to a remote NAS using the private cloud function provided by the NAS vendor.

Retrospect and the LTO-6 tape drive are the newest additions to my backup routine. Retrospect and the tape drive are installed on my PC, and agent software is installed on the Linux server and my wife's PC. After an initial full backup of data on all three systems, nightly incremental backups are run for a month or until the tape is full. At that point the tape is replaced with a blank tape and the process repeats. I use Retrospect's *Duplicate* function, which copies data in native format, to archive data to an LTFS-formatted tape on a yearly basis.

As I said, this is probably overkill. Remember: 3-2-1 is a minimum, not a maximum.