Backup, Backup and Backup

We often admonish users to backup their systems. Many of the calls for help that we get on the air could have been avoided if the listeners had a practical way of restoring the original information on their systems.

But this is only possible if the information was properly saved.

What we all want is a magic wand that can reset a system to its original state. When we say “backup,” there are really three parts to this process.

Backup – a copy of the information
Retention – an organized method to make sure the information is available
Recovery – restoration of original information onto a system

This document gives an overview of how you can backup and protect your system. We also discuss the exact tools and the “how to” procedures that you’ll need.

There are two types of basic information stored on your system. The first type is the operating system — such as Windows — and the application programs. In general, it is not necessary to backup this information because one can go back to the original media — CDs or floppies — and reinstall the programs or reconstruct the operating environment. Most brand name manufacturers provide a recovery CD that lets you easily restore the system to “as new” state. After doing this type of restoration, you will have to reinstall all other programs and patches that you had installed since the system was new.

The other type of information stored on your system is the user data that has been created or collected since the system was new. This includes documents — such as reports or emails — and customized features — such as Windows configuration settings — that you have created or downloaded or changed in the course of using the computer. This information needs to be saved so that it is also can be restored in the event of data loss or hardware malfunction. Unfortunately, identifying all the information in this category. Different versions of Windows stores data in different folders. For example, it is fairly clear that you will want to save the contents of the “My Documents” folder, but knowing what to save in order to retain your display settings or email account details may not be as easy to determine. Here are a few examples of information that should be saved on a regular basis.

Web-browser bookmarks or favorites
E-mail messages
Syncrhonized PDA files
Data in folders from specific programs
User-created data such as documents and spreadsheets

If you already have a complete backup of your system, it is only necessary to save data or files that were modified or created since the last backup. However, this can be very inefficient since you will have to keep track of the bits and pieces of these “incremental” saves. You can save time when creating the backup, but when it comes time to restore lost data, the process can be made much more complicated if you have a large series of incremental backups. As a result, it is easier for the average personal computer user to backup the entire system’s data because everything is kept in one place. In a corporate environment, incremental saves may be a better strategy simply because of the large volume of data that must be stored.

However, full system backups are complex and difficult, especially with the large hard drives that many users have today. It’s not easy to backup to multiple pieces of media, and it’s not always easy to restore from them. It’s not easy to find the time for a full backup, because most can’t do it unattended since multiple pieces of media will be required. It is better to perform a full backup on a regular basis — quarterly, monthly — and then perform incremental backups on a more frequent basis depending on how critical the data is.

Backup data by category.

You should backup working files as you create and use them. For example, create backup copies of reports and spreadsheets when you first create them and again when you finish working with them for the day. If you keep all these files in a folder that is easy to find — such as My Documents — it can make it easier to backup these files. Other work files — such as financial data in Quicken — can often be backed up using a command within the program itself.

BACKUP MEDIA

What type of media should you use for backup?

The answer all depends on the volume of data and what you have available on your system. Your system may have one or more of the following types of removable, writable media.

MEDIA	Capacity	Number required per 10 GB data
3.5-inch floppy	1.44 MB	6,944
ZIP disc	100 to 750 MB	100 to 14
CD-R or CD-RW	780 MB	15
Tape storage	2 GB to 250 GB	5 to 1
DVD writable or rewritable	4.7 GB	3

Other possible storage media include removable hard disk drives, USB storage devices (keychain, pen, or fob storage), and memory cards such as SmartMedia and PC Cards.

Retention

Is it necessary to have more than one copy of a backup?

The answer is definitely yes. If you only have one backup device, and you suffer a system crash while writing to it, you risk losing everything, leaving you with no data to restore. We recommend that you have as many as five different copies of your backup, and rotate them. For example, if you're making daily backups, have one copy for each day in the week.

In addition to multiple copies, good retention practices include storing at least one backup copy off site, away from your computer. Some people keep one copy at work, one at home, and one in their car. For example rotate the multiple backups off-site. — i.e. on Monday night, after you run the back up, take the previous Thursday's tape home, and put the previous Friday's in the car. You now have the latest back up at the office, the next oldest in the car, and the oldest at home. In case of multiple disasters you have limited your loss.

Unfortunately many computer users can't be bothered with backing up because it DOES take time and effort. What may be minimal effort for some is "too much" for the student. Doing a backup without having concerns for retention is only half a loaf. I have found that buying a compact flash card with appropriate storage and a PCMCIA/CF adapter (street price $10) answers part of the question of backup. When the thesis is due at the end of the semester, it is nice to know that this minimal cost of insurance is protecting today's' higher education investment.

Backup is also a way of offloading from the system infrequently accessed data on the hard drive thus freeing up hard disk storage. The classic example of this is the annual income tax return. Once done, it doesn't have to sit on the computer and, indeed, for security and safety you might want to store it separately.

Backup is useful for those who may have accidentally erased information.

The last phase is recovery.

In order for a backup to be of any value, you must be able to restore it to your system. This is the process of restoring "saved" data. This may sound silly, but all too often, we know to save or backup our data but we really never test the restoration of data. It is important to know if the save is complete. Without this, you can't have confidence that the recovery data will be there when you need it. When necessary, you want to know how to reload the system properly.

This is very much like having a spare tire in the trunk. That is a good start. But you must make sure that the spare is inflated and you can operate the jack. It's one thing to have a spare, but another thing if you don't know how to change a tire, or it has no air in it. (And as with a tire, if you decide to have a professional do the work, that is, restore your data for you, it can be an expensive proposition.)

Proper backup coupled with Windows "System Recovery", or Symantec's' "GOBACK", or a comparable program or system is must do today.

Organization of Data

The backup or copying of files can be a very simple process or it can be a very tedious and complex process. How the data is organized is the determining factor. The time you spend on planning the organization of data will impact the efficiency of the backup process.

If you are using Windows XP, by default your primary data is stored under your user profile name. You can locate this easily by the following:

My Computer ->C:/(Drive)->Documents and Settings->(userid)

Where (userid) is the currently logged on user.

Two folders of note under <username> are "My Documents" and "Favorites". If you are using Microsoft Office or other Microsoft applications, all user documents are saved in the default location "My Documents". However many third party applications save their data in folders under the "Program Files" folder.

My Computer -> C: (Drive)->Program Files -> <3rd Party Application>

If the option is available to custom install these applications, then directing the application work area to "My Documents" will greatly simplify the backup process.

There are two aspects to data protection. The first is the saving of data or "Backup". The second is the "Recovery" or restoration of the original data.

When information is copied to an external medium, the backing up of a folder to external storage is straightforward. But when a "Recovery" or restoration of data is required, you need to redirect the saved information to the correct working folders.

Fortunately for most people, the need to effect the recovery phase is infrequent. But the down side of this is that when you need to do a recovery, you won't remember where working folders are located. That is the primary reason for storing working data under one roof or folder.

There are three basic kinds of data. They are:

1. Application Programs (e.g. Word, WordPerfect, Excel)
2. Regular or Dynamic Data (word processing documents, and spreadsheets) that require daily backup
3. Relatively Static Data (e.g., Palm Pilot used for contact list)

1) Application programs need not be saved.

The Windows operating system falls in this category. It can be easily regenerated with a new or upgrade installation from the distribution CD or diskette. One of the major obstacles many have encountered in reinstalling Windows is the tedious step of reinstalling the “patches”. In the case of Windows older than Windows 98, updates are no longer available on the Microsoft website. Support for Windows 98 has been scheduled to be withdrawn at the end of 2003.

If possible, save all patches used for a Windows update. This eliminates the question of availability as well as the time it takes to download. Some of these patches are very large in size. If you are using a dialup modem for Internet access, some of these downloads go beyond one hour.

Microsoft offers (at no charge) a CD with all the updates for Windows XP, Windows Me, Windows 2000, Windows 98, and Windows 98 Second Edition (SE) and they claim it takes 2-4 weeks for delivery if you order it from http://www.microsoft.com/security/protect/cd/order.asp

For Windows 98, the updates are located at
http://www.microsoft.com/windows98/downloads/corporate.asp
We recommend you download them and burn them onto a CD or some other removable storage.

When the subject of backup is brought up, some people think it is necessary to backup the entire hard disk drive. In general, the information that needs to be saved is much smaller than the total usage of the hard disk drive. First of all, your hard disk drive should not be filled to maximum capacity. If it were filled, you would not be able to operate properly. My rule of thumb is not to exceed 75% of maximum capacity. If I am consistently running the hard disk at more than 75% capacity, then it is time to consider upgrading the storage device.

2) Dynamic data changes frequently. An example of this is a working document in Microsoft Word. Other examples of data that are apt to change daily is the message save area of your e-mail reader, your appointment calendar, or the tracking of your investment folio.

3) Finally there is static data. These files also need to be saved but it is not necessary to do a daily backup since they change infrequently for the most part. An example of this may be updates to information on your PDA. This type of data should be backed up whenever it changes.

The reason for distinguishing between dynamic and relatively static data impacts on the storage medium as well as the time it takes to do a backup. If you are working with an existing system, you should review how the data has been organized. If it is possible to reorganize the data, then do so by all means. Otherwise I would keep notes to remind me how the hard disk has been organized. It is obvious that for those who have more than one computer, relying on memory is less than ideal.

In a small office environment, consistency and standardization greatly simplifies the backup process. It is important to note that files need to be closed before they are copied. In the case of a small office, all users of a network application must be “Logged Out”.

Organize your data and the process of backup becomes easy.

Peripherals and Media for Backup

We recommend that removable media or network storage be used to backup information. Using internal storage such as a separate hard disk partition or even a separate hard disk drive is not prudent. If the backup partition resides on the same physical drive as the primary partition, then a physical hard disk failure can render all partitions with information on the hard disk drive lost.

Similarly, a fixed disk controller failure can cause loss of data integrity on all drives connected to it. Finally a fire can destroy all components inside a system unit. I can attest to this since I had the misfortune of a major house fire during which all my computers and drives were affected. Only off-site backup survived unscathed.

There is also offsite online storage from many vendors. Some use Yahoo Briefcase for offsite storage, but others have been burned when some online suppliers went out of business with little or no notice.

Most enterprise or corporate systems use at least three cycles of data retention. In a three cycle arrangement, backup is written over the media created three days earlier (e.g. Thursday backup is written over Monday media). On a given Thursday, if recovery is required, the Wednesday information is available. If for any reason there is unreadable information on the Wednesday media, you can use the Tuesday backup. If in the creation of the Thursday backup, a problem was encountered, the Wednesday (immediate) and the Tuesday (one day older) information is still available. In the worst case, there is loss of up to two days of

work. Businesses normally file their paperwork a few days after processing. In this way, the source information is still available for data re-entry.

However we normally recommend that, for personal use, you have media marked Monday, Tuesday, Wednesday, Thursday, and Friday. This simplifies the need to understand or to follow the proper rotation of media. And most individuals do not have so much data to backup that the extra media would be a hardship.

Incremental versus Full Save

When you are running a backup, one technique is just to save or copy files that have been modified. This is referred to as an incremental save.

If at the beginning of the work week, you copy or save only data files that have been modified since the last time you ran a backup, this will cut down on the time to backup a system. Additionally it will also reduce the amount of storage media needed for the process. Then once a week, at the end of the week, a full save of everything is performed. The advantage is reduction in time and media. The disadvantage however is that you will be required to go back to the beginning of the week to do a day by day recovery if the entire system had to be restored. In a corporate or enterprise environment, the operations staff is trained to do this. For the average personal computer user, this is confusing and time consuming; so we recommend a full save for backup purposes

Backup Media

The exact media you use is largely dictated by the amount of information you have identified for backup. The following are the commonly available removable media storage for the personal computer of today:

Floppy Diskette

Reliable; very low cost; very limited capacity; media universally available; slow backup time; moderate speed for recovery of data file(s)

Main advantage – handling, cost, availability

Main disadvantage – very limited storage capacity; limited shelf life; end of product life.

ZIP Drive

Reliable; reasonably low cost; reasonable capacity; end of product life cycle; reasonably fast backup; reasonably fast recovery of data file(s)

Main advantage – super capacity floppy diskette

Main disadvantage – end of product life.

CD-R or RW / DVD

Reliability dependent on handling and manufacturing quality control; very low cost; high capacity; media universally available; fast backup time; fast recovery of data file(s)
Main advantage – high capacity, cost, availability
Main disadvantage – delicate handling of media

Tape Storage – serial operation

Reliable; low cost; very high capacity; limited media availability;
Very fast backup; very slow recovery of file(s)
Main advantage – very high capacities
Main disadvantage – serial operation for recovery of file(s)

The following are factors that will determine what media you will use:

1. Similar to configuring minimum hard disk storage, plan on a backup storage medium that is at least 25% greater in capacity than you actual estimate you need. This will enable you to handle any spikes in usage. Otherwise you will encounter incompletion of backup. This is more critical when backup of data is done in unattended mode. An example of this is the backup of the server after working hours in a small office environment.

2. The time it takes to save the data. In an attended operation, you will be impatient to have it completed as soon as possible. In an unattended operation, it is still a critical factor. If the backup process was incomplete, it will eat into regular work day hours unless you want to chance a “no backup” for the day before. In a small business environment, the backup must be completed before the start of work the next day.

3. The cost of media is not trivial. Anyone who uses the Jazz drive knows that the cost of backup media for the week can be in the hundreds of dollars. Media does not last forever. It is recommended that the media be replaced annually—before a failure. The weekly or daily use of media will wear down the media.

The handling of DVDs and CDs is a different problem. One must be very careful that fingerprints not come in contact with the recording surface. The oils from one’s body can be destructive.

4. One should be aware of the ready availability of media. If you need a quick replacement, it may not be sold at the local computer or stationary outlet. An example are the 100 Mb ZIP cartridges. They are nearing the end of a long product cycle. It is very important that the peripheral and media can easily be purchased. If you had to use the backup on a new or “borrowed” system, the storage peripheral must either exist or be readily available for purchase. Otherwise the backup is not really a usable backup.

“Quickie” Backup

This is another category of backup which falls outside of the general rules we have set, above.

First is “System Restore” in Windows XP and Windows ME. This enables you to easily return to a prior point in time matching a “snapshot” of the drive which you have taken. ConfigSafe from Imaginelan.com does a similar job. Symantec’s GoBack is a third party system that has the additional ability to enable the user to selectively “roll back” a specific file to an earlier version. This keeps you from losing a current file if you do have to roll back your entire system.

These systems when used with standardized backup procedures will enhance the overall integrity of your data.

Equipment Transition

Changes in Technology

The one constant we have in technology is that it will constantly change.

Some products have a short product life cycle and other products don’t seem to go away. The 3.5” floppy diskette drive is still common equipment on a desktop unit, though some product lines are shipping without one being standard. It was originally introduced by Apple in the early 80’s and adopted by IBM with the introduction of the PS/2.

During that period, tape drive and tape media technology has changed many times over. But as new removable storage replaces the old, it is necessary to ensure that permanently saved data is converted.

When technology changes, there are three things to consider:

- Converting data storage formats with changeover in equipment.
- Coexistence of unlike storage media
- Review when data was last created to ensure data integrity

1. When new equipment replaces old, it is necessary to be able to convert the old data on some common system. An alternative is to retain the old system on a local area network. This will enable access and transfer of archived data. There are outside conversion services. You may want to consider this based on the volume and cost effectiveness of maintaining legacy technology on a new system. Whatever you choose to do, you must be able to recover the data or else it is not a backup.

2. A more common scenario is the mixing of the new with the old. Many small businesses use the life cycle of the equipment to make upgrades. A commonly accepted product life cycle of a desktop unit is three years. In this case, there is a one-third replacement of systems each year. Those who are power PC users can always have the latest equipment while their systems are in turn trickled down the organization. In this way there is a gradual integration of old with the new. Otherwise all the systems will grow old together making each system upgrade a major conversion. You would have to convert all data on old storage media at one time. Using old and new on a local area network, it is possible to have a paced schedule of conversion.

3. Another problem that many have experienced is the inability to read data originally written on a floppy or ZIP drive years ago. Data written on magnetic media is not really permanent. Magnetic data can and will degrade. Many companies have found magnetic recording surface on tape media greater than 10 years old flakes when read which in turn causes damage to magnetic tape drives. It is necessary that you periodically take inventory of saved or backup data and review when it was last created. Take the time to “refresh” the data by copying it onto like or equivalent media.

Media Expense

As we had noted, some information is regularly replaced by updated data. This requires new backups to be made regularly. But you will also have data that never changes, for example, digital photographs.

Most likely, the preferred storage media for this category is CD-R or DVD-R (see the description of the various DVD formats at the end of this article). Although manufacturers have claimed a life of 70 years or more for CD media, recently, a report has been published that some of this media is degrading in a very short time—in some cases a matter of a few years. We do not consider the study to be scientific because the bulk of the test media in the test were white label or no name CD-R’s. These may or may not have been rejects that have been put into the marketplace. However, until there is certified testing which includes brand name labels, we do recommend the use of brand name labeled CD and DVD media for storing permanent data; frequent testing for failure; and recopying at regular intervals for data which is considered permanent archives.

Considering the importance of the permanent data, the extra cost is minimal. It is foolhardy to save valuable data onto the least cost medium. Remember CD and DVD requires careful handling. Since there is also the potential for dropped data bits, make a duplicate copy. The media expense should not be a factor.

There has been some controversy over writing directly on the surface of the CD/DVD media with a felt-tipped marker. While I have not had any difficulty

caused by this, prudence suggests that you use only pens certified for this purpose.

If you do find that there has been a failure of the CD or DVD, there is a class of software which can be used to try to rescue the data.

DVD X Rescue and CD X Rescue from 321 Studios provide users with the ability to fix a favorite movie that jumps, an audio CD that skips, a photo disc with treasured photos that can no longer be accessed, and other personal and business data that doesn't read.

Storage

How one stores media impacts the overall integrity of the data. First of all, as we mentioned above, a full backup should be stored off-site in case of disaster.

Media should be kept away from temperature extremes. Locate the media, if possible, in an area of low humidity. Ideally one should save data in a separate and secure area. If possible, store in a fireproof box. Or have it stored with a trusted person. When data is transported from one location to another, allow time for the media to acclimate to the operating environment. An example of this is media brought inside after being exposed to cold weather.

In a small office environment, one person is normally charged with taking home the oldest backup and bringing back to work the oldest media to work the next day to be used for backup.

Media Cleanup

When media used for backup is cycled out of the system, it is important to remember that there is "live" data residing on it. You want to make sure no one is able to review discarded "sensitive" information. A bulk eraser may be used to "zap" the magnetic storage media. Be careful when in use that it is not near "live" media. The erasure is not absolutely foolproof. But under normal circumstances, it is a practical method. If you are dealing with sensitive material, there are programs available which will, after erasure, completely overwrite the data with bogus data, such as those which can be found at http://download.com.

Also, remove or obliterate the labels. Otherwise it is a walking advertisement. For optical media, you should physically destroy the disk (there are CD-shredders, but folding, and scratching the surface should suffice.

Note: DVD formats

DVD-R
(DVD-Recordable) A write-once (read only) DVD disk for both movies and data endorsed by the DVD Forum. DVD-Rs are often called "DVD Minus Rs" or "DVD Dash Rs" to distinguish them from the competing "Plus R" format (see below). DVD-Rs are the DVD counterpart to CD-Rs and use the same dye-layer recording technology to "burn" the disc. Pioneer was the first to introduce DVD-R drives, which recorded 3.95GB. By 2000, the capacity was increased to the industry standard 4.7GB.
In 2000, DVD-R was split into two types to deal with copy protection. The original DVD-R, which uses a 650 nm recording wavelength, was dubbed "DVD-R for Authoring." A different format with copy protection that records at 635 nm is called "DVD-R for General." Although DVD-R(a) and DVD-R(g) can read each other's format, they cannot write each other's format.
DVD-R machines (DVD burners) cost as much as $17,000 in their first incarnations back in 1997, but dropped to under $200 in 2003.
DVD+R (DVD+Recordable) A write-once (read only) version of the DVD+RW optical disk from the DVD+RW Alliance. DVD+Rs hold up to 4.7GB of data per side and can be read by DVD-Video players and computer DVD-ROM drives.

DVD+RW
(DVD+Read Write) A rewritable (re-recordable) DVD disk for both movies and data from the DVD+RW Alliance. DVD+RW media can be read on DVD-Video players and computer DVD-ROM drives. Using phase change technology, the first DVD+RW disks held 3GB per side, but were later increased to the industry standard 4.7GB. A double sided disk holds 9.4GB. DVD+RW supports both the CLV and CAV recording formats, the latter providing more uniform random access for interactive data applications.

Footnote:
What are the three most important words when you use a personal computer?
Work Smart. Backup!