How File Recovery Works

Introduction

Mobile / cell phones, digital cameras, MP3 players, laptops, games consoles, GPS sat-nav – it seems that on almost every device we use, there is digital information stored on it. The internet, electronic gadgets and computers seem to be involved in almost every aspect of our everyday lives.

With all of this data we are finding that some of the most important aspects of our day-to-day lives – our music, our photos, our home videos, our contacts, our emails, our financial information, etc, are being held on memory cards little bigger (and often smaller) than a coin in your pocket.

Let's take cameras as an example. Not too long ago we pack our camera for our holiday / vacation, load it up with film and click away (but not too much as you didn't want to waste that expensive film). Taking videos was left to those people with more money, more muscles to carry around the equipment, and a seeming predilection to filming almost "anything that moved" and often things that didn't.

Once back from our holidays, we'd drop off the film to get developed at a local shop. We'd return several days later in anticipation to collect our precious memories and re-live the holiday all over again. Of course there was always that one picture that we anticipated the most, the one that we knew was perfect – the one of the beautiful sunset, the holiday romance, your child's moment of utter happiness, the amazing view. Often, we'd find an unfocused shot of someone's foot or with someone's head cut off from the nose up. (Normally after you'd stopped someone and asked if they would mind taking a photo….)

After sharing the holiday stories illustrated by the photos with your closest family and friends, you'd then spend the time sticking them into photo albums to be placed on a shelf fondly next to other past holiday and family moments captured.

The digital camera has changed all of that. With ever-increasing memory cards, we have all become snap-happy. We can immediately see if the moment is captured to our satisfaction and re-shoot it then and there, as needed.

Initially, everyone would still print off their photos and stick them in photo albums. But as time passed, our options changed and our cherished photos and videos now tend to stay on computer hard drives, DVDs, or websites like Flickr or Picasa.

With so much of our personal information and memories now on hard drives and memory cards, there's an often unthought-of risk. What happens if something goes wrong? You click "Delete" by mistake, your camera suddenly can't seem to read the memory card any more or while your child was unknowingly playing with your camera, they accidentally formatted its memory card. Ultimately, all hard-drives (being mechanical) will fail... But you can always restore to that backup you've taken. You did take a backup and keep it up-to-date, right?

This article aims to provide a high-level technical guide to file recovery. It will help you understand what is possible should you lose some or all of your files, and also why it is not possible to recover files in some situations.

But, before we begin with how to recover your files, let's start by looking at what's going on in your computer's hard-drive and on those memory cards.

Drives: The Structure of Hard-Disks, CDs, DVDs, and Memory Cards

Let's begin by defining some basic terms. The diagram below shows how the drive logically holds data:

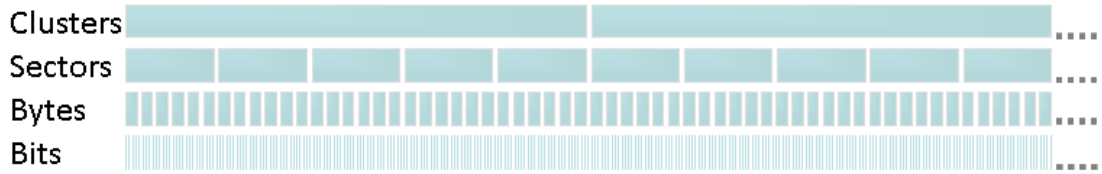


Figure 1 – Data Structure

The lowest level of data in any computer system is a bit (**b**inary dig**it**). A bit can have only one of two values: 1 (on) or 0 (off). To make bits more useful, they are grouped together into a "logical construct" called a byte. A byte consists of 8 bits.

So far, so good. You will often hear the terms bit and byte used frequently. For example, your internet connection is usually defined in megabits (a "mega" in non-IT circles is generally considered to be a million while a "giga" is a billion). So, your internet connection might be defined as being 4 Mb (4 million bits) per second – note the small 'b' for 'bit' in Mb.

Memory cards however, tend to be defined in bytes. A 100 MB (MegaByte – note the capital 'B' for 'Byte') card has 800 Megabits (8 x 100). That internet connection doesn't sound so quite impressive any more does it?

Note: When talking about data and data throughput there is an accurate IT definition (based on binary – remember everything digital is ultimately a 1 or 0) and the "marketing" definition which is considered easier to understand. For example: The IT definition of a 1 MByte drive is 1,048,576 bytes but the marketing definition rounds this down to 1,000,000 bytes.

As can be seen in the diagram above, a drive has two more groupings. A number of bytes are grouped into what's called a sector, and a number of sectors is grouped into a cluster. The number of bytes in a sector and the number of sectors in a cluster varies from drive to drive.

Why are sectors and clusters needed? They exist to make disk manageable more efficient – a little like page numbers in a book. Trying to refer to something in a book as being 42,491 words from the beginning would be a little tedious and time-consuming to find! Using chapters, page numbers and paragraph numbers (i.e. chapter 3, page 75, 3rd paragraph) makes the information in the book a lot easier to find.

Logical Drive Structure

A drive consists of three main sections as shown in the following diagram:

Figure 2 - Generic Drive Structure

Internally a drive can support many different format structures such as FAT, FAT32, NTFS, CDFS, etc. Although they differ in their details their basic operation is the same:

> The **Boot Sector** is an area at the beginning of the drive that defines its structure and tells you everything you need to know about ***how*** to read the drive.
>
> The **Index** section provides information about the files and folders that exist on the drive. It is here that the file name, type, size, status, folder, etc are held as well as the crucial information of where the file contents are located on the drive.
>
> The **Data** section of the drive holds the actual contents of the individual files.

A file can only start at the beginning of a cluster. Any space not used at the end of the last cluster is wasted.

As mentioned above, a drive is split into clusters, clusters are split into sectors, sectors into bytes, and bytes into bits. This means that a cluster has a fixed size. So what happens if a file's size is larger than the size of a cluster?

When a file is larger than the size of a cluster then clusters are linked together to form a c**luster-chain**.

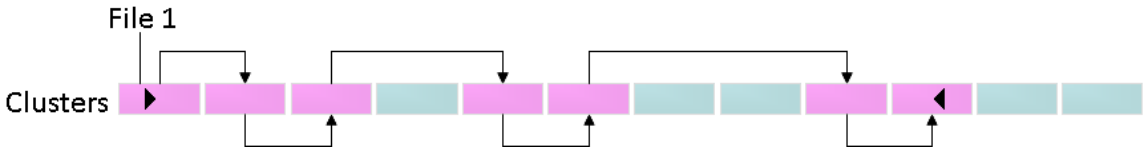


Figure 3 – Example Cluster-Chain

Information about this cluster-chain is held in the Index. Note also that the clusters don't have to be contiguous (i.e. sequentially next to one another) In fact, large files stored on frequently-used drives often aren't.

To read a file, a program looks up its entry in the Index, and then reads the data held in the clusters of its cluster-chain.

Fragmentation

As we have seen above, a file's clusters don't have to be contiguous but can be spread across the drive's clusters. A file whose clusters are distributed in this manner is ***fragmented***. An example of this can be seen in the following diagram.

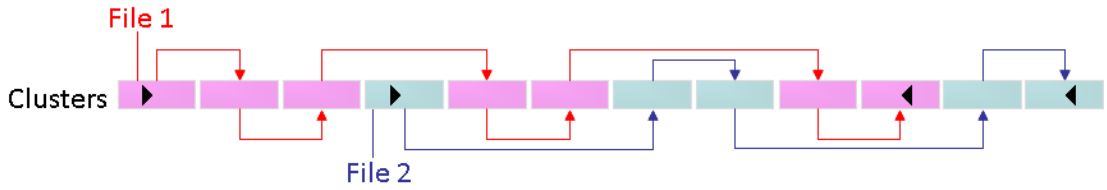


Figure 4 - Fragmented Files

The following diagram shows files that are not fragmented.

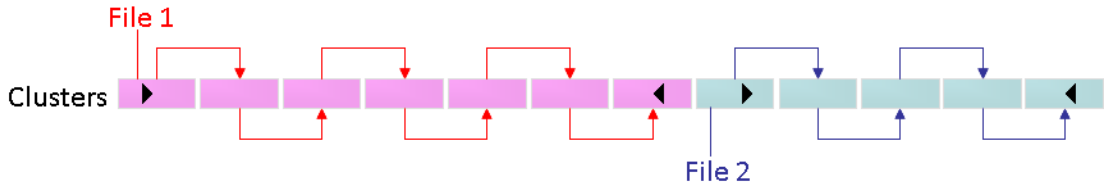


Figure 5 – Non-fragmented Files

Why is fragmentation important? It is important for two main reasons: speed and recovery potential. Imagine trying to read a book where the pages are not sequential. You know where the next page is but you have to find it before you can read it. This would impact your reading speed because of the additional time to find the pages. It is the same with computers reading fragmented files. We will see later how recovery potential is impacted by fragmentation.

Deleting and Formatting

When a file is deleted by the common Delete function, the deleted file's data is actually left untouched on the drive and only its Index is flagged as deleted. That is, the file's associated data space is marked as available for use / overwriting.

Similarly, when a drive is formatted there is a good chance of the data still being available. The "Quick" format option actually only rebuilds the Boot Sector and the Index. The data section itself is left intact. This is not only faster, but provides an opportunity to recover many of the files on the drive.

File Recovery Techniques

So how does file recovery work? Well, there are two main methods.

Undelete Scan & Recovery Method

File Undelete works by reading the drive's Index and identifying those file entries that have been marked as deleted.

From the index, a program can determine a file's name, its logical folder location, its size, and its cluster-chain. Using this information, the file recovery program can *attempt* to rebuild the file. Why "attempt"? The information stored on drives changes frequently and often without our knowledge as the device controlling the drive (our phone, camera, computer, etc) will often be performing tasks ***behind the scenes***. These actions can sometimes lead to new (temporary) files being saved and deleted and these files could be

re-using the space now set to being available by files that have been “deleted”. What this means is that although we may be able to find the beginning of a file’s cluster-chain, the data it references may no longer be valid. That is, part of the cluster-chain’s contents may be overwritten.

Undelete file recovery has several advantages:

- **Speed** – because it initially only reads the Index portion of the drive it can be quite fast.
- **Details** – As the Index holds details about the file’s name, its size, its logical folder location, and its type, this enables the recovery program to recover the file with its original details, known as attributes.
- **Generic** – When a file is found using the drive’s Index, it is simply treated as a group of standard attributes and some associated data – there is no ***meaning*** to the information. What this means is that there does not have to be any understanding of the underlying file type to be able to recover it. Why this is important will become clearer later.
- **Fragmentation is supported** – Having the Index means that we have the file’s cluster-chain and are therefore able to piece together the different portions of the file.

The disadvantages are:

- **Deletion Only** – If you have formatted your drive or the index has become corrupt then the Index, which is fundamental to this recovery method, has been re-created and the information it held has been permanently lost.

The Undelete technique is useful when you have accidentally deleted a file and not continued saving much, if any, data to the drive but what if you have formatted your drive or the drive´s index has become corrupt (i.e. files just suddenly disappear)? What other techniques exist?

Low Scan & Recovery Method

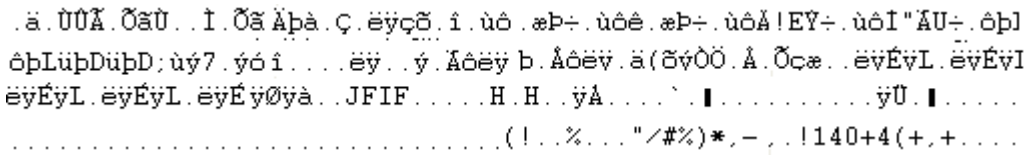
To understand how this method works, it is important to know that many of the files that we use have something known as a ***File Signature***. This signature is a sequence of characters that occur in a predefined part of the file. Usually, this file signature is found at or near the beginning of the file.

Continuing with our book analogy, this is similar to trying to find a chapter in a book with out the index. You know that a chapter will begin (as an example) with the word “Chapter”, so you don't need the index to find the beginning of a chapter though it would have been quicker had you had the index to look up the page number.

Once the user selects the file type(s) (i.e. file signatures) that they wish to find, the file recovery tool will scan the drive from beginning to end (like a record, or tape) looking for the

desired file signatures. It assumes nothing about the drive's structure[1] or its index, as these might have changed or been re-created since the files were lost.

The following diagram gives an example of how a Low Scan sees a drive.



If you look carefully, there is the beginning of a jpg file in the data shown. (ÿØÿà)

Although many files have file signatures there can be some issues associated with them:

- **Many files have signatures but no footers** – A ***Footer*** is similar to a File Signature except it occurs at the end of the file. Footers, unsurprisingly, are useful for knowing where the file ends. Remember, the Low Scan has no access to an index and so knows nothing about the file's size or where the fragmented parts are located.

  So how does the Low scan know when it has reached the end of a file? The simple answer is that it doesn't so it makes an assumption. If it finds the signature of another file that it is looking for then the previous file is probably finished. This method works remarkably well but can sometimes lead to a file that was originally several KiloBytes in size becoming several MegaBytes after the recovery. Although this sounds disastrous, this isn't as bad as it initially seems as opening the recovered file in the application that created it and re-saving usually "loses" the additional "junk" bytes.

  It is possible to read the size of some file types from the file itself but this means building specific functionality into the recovery application, so tends to be implemented only for popular file types where the information is available (such as jpg files) and more advanced file recovery tools like Media Investigator.

- **Files can't be fragmented** – This is where a file's data is distributed in different areas of the drive. Remember that the cluster-chain where these areas are is held in the Index.

  As stated earlier, the ***Low Scan*** only reads a drive's data section from beginning to end and knows nothing about the Index. What this means is that if a file is fragmented then it will not be possible to fully recover it as the location of its clusters is unknown. (If we knew them then we could use the faster ***Undelete Scan***).

- **File details are unknown** – Again, it is the Index that holds details like file name, logical folder location, size, creation date, etc. The Low Scan knows the type (because of the file signature) but that is it.

- **Speed** – Or more accurately, lack of speed. The Low Scan reads the drive from beginning to end. A 2GB memory card has 2,147,483,648 bytes and an 80GB hard drive 85,899,345,920 bytes – almost 86 billion bytes! Although recovery programs try to optimize their code, there is no quick way of reading, and more importantly interpreting, that much data. When you consider that an average file signature

[1] Generally – there are some exceptions.

consists of only 6 bytes then it really is like looking for the proverbial needle in a haystack.

- **False positives** – File signatures are basically a pattern of data, albeit a small pattern. That pattern could randomly occur in other files' data or even in the same file. Special algorithms are employed by advanced recovery programs like Media Investigator to reduce the impact of these so called ***false positives.***
- **File type specific** – When searching for specific patterns you must know the pattern that you are looking for. This means that the recovery program must have been supplied with the data patterns (file signatures) that you need in advance.

There seem to be quite a few negatives but what are the positives? The positive is that the Low Scan method is your last hope. It might not find everything but there is a good chance that it will find something – possibly your most important lost file.

This leads us to ask the question: What can be done to maximize our chances of recovering our files if they are lost?

Maximizing Your File Recovery Chances

Mistakes happen. Files get lost, accidentally deleted, formatted and even corrupted. So what can we do to maximize our chances of recovering our files if things do go wrong?

The most important rule to maximize your odds of a 100% successful file recovery is: **Stop what you are doing!**

If your are using a computer then stop what you are doing and start the recovery process as soon as possible. The faster you act, the more likely that the faster Undelete scan will find your file(s).

If you're using a camera, don't take any more photos or videos.

If you are using an mp3 player, don't add any more music.

Just as importantly, don't save your chosen file recovery program to the drive which contains your lost files.

The above may all seem quite obvious but there are also things you can do ***before*** losing your files that will help maximise the chances of a successful recovery should you need it.

Computer and Drive Recovery Precautionary Actions

- **Regularly de-fragment your drives** – If a file is not fragmented then the chances of recovery increase dramatically plus, a less fragmented drive will speed up your computer!

  Most computers come with some defragmentation software pre-installed. (In Windows, see: Start/All Programs/Accessories/System Tools/Disk Defragmenter or you can right-click on the drive, select Properties and then choose the Tools tab.)

Digital Camera Recovery Precautionary Actions

- **Don't use your camera to delete photos or videos** – The beauty of digital cameras is that you can see your photo or video as soon as it is taken and then, if it's not as good as you'd like, you can delete it. To maximise your file recovery chances should the worst happen however, it is best to not delete them.

  Why? It due to fragmentation. If you delete files, you will leave data "gaps" on the drive which the camera will try to fill when you take more photos or videos. The chances are that new files will be fragmented as they re-use the drive space occupied by the deleted photos. Fragmented files reduce your chances of a successful recovery, should you need it.

- **User smaller memory cards** – The larger the drive, the longer between taking the photo and the saving it to your computer. The longer the time, the greater the chance of losing the picture. It's also a case of "all your eggs in one basket".

  So instead of buying one 4GB (or larger) memory card, you may be better off buying two 2GB cards or even four 1GB cards from a recovery perspective. Any recovery needed will also be faster since the card will be smaller. With the price of memory cards dropping in recent years, it is now a much more affordable measure to have several extra cards that you can use until they become full.

  Once they are full, you can transfer only the photos you want to your computer and delete the rest or format the memory card.

- **Format your cards before using them** – This is best done just after you have transferred your files / photos / videos to your computer rather than just before you start using it in your camera. Otherwise you risk accidentally formatting a card that you have already used but not yet transferred the files from.

  Why should you format memory cards instead of just deleting files? It comes down to fragmentation again. By formatting the card, you start off with a ***clean slate***. You may still have remnants of old data on the memory card but the Index will be cleared and the drive won't have any ***gaps*** to fill in.

- **Use "quick" formats** – There are different ways of formatting a drive. Some overwrite the whole drive, Boot Sector, Index, and Data. Others just rebuild the Boot Sector and Index. To increase your chances of file recovery (should you need it), you need data to recover on the drive and so should use any "quick" format options instead of "Low", "Full" or "Zero" format options since the "quick" format option will actually leave the data intact.

File Recovery Software – How to Choose

You now understand the technical approaches to file recovery and some pre-emptive measures you can take but if an accident does happen and you lose your files, what should you look for when searching for file recovery software?

- **One product** – As covered earlier, file recovery programs use some standard approaches to their recovery algorithms and processes to recover files. Although some file types are more difficult to recover than others, the general approach remains the same.

  Several of the companies that provide file recovery tools offer a range of file recovery software. Some work with memory cards, others with hard drives. Some with emails, others for photography.

  It is far more cost-effective for the consumer to buy one program that supports all file and media / drive types. Plus, you'll have used less disk space by installing only one file recovery program instead of several.

- **Multiple file types** – Connected to the point above, the product you purchase should support as many different file types in as many different interest areas (i.e. multi-media, business software, photography, music, etc) as possible but of course must support the file type(s) you need or could need to recover in the event of file loss.

- **Previews and Information** – Before spending any money it would obviously be helpful if you could see the files that you are recovering before actually saving them. Technically, this isn't quite as straightforward as it may seem as there are so many different file types to cater for but the basic jpg, gif, and bmp image files should have a preview function which shows you a preview of the actual image that will be recovered – even if it's only a partial file that will be recovered.

- **Guarantee** – File recovery itself is never guaranteed due to the multitude of factors which affect the likelihood of success which means that, despite best efforts, some files will be unrecoverable.

The problem is that you often don't know if the files recover fully until you have saved the files and try to use the. But to save them, you need to pay for the file recovery software. But you don't want to buy the software unless you know it will recover your files… It's a bit of a vicious circle, isn't it?

A company which backs its file recovery technology will offer a money back guarantee if you are unable to successfully recover your files. Naturally, once you have recovered some files, this guarantee is void but it should give you peace of mind for the first recovery you perform.

- **Price** – Price is important to any purchase you make and file recovery is no different. File recovery programs range from free to several hundreds (or sometimes thousands) of dollars for the most sophisticated recovery tools and, like almost any product or service, there is usually a trade-off between the price of the software and the value you place on your files.

Like most products, file recovery software takes time, effort and skill to create and support personnel to provide the best support. All of these factors cost money which a company needs to price into its products. Otherwise, the company won't be able to offer products for long.

As suggested above, it is usually more cost-effective to buy a single file recovery tool rather than a portfolio of supposedly “specialized” file recovery tools.

As recovery is often down to a “one-off” mistake and you may not want to actually buy a licence to use the software indefinitely, some vendors also offer a low-price “Single Recovery” fee (effectively, a pay-per-use licence) so you don’t have to purchase an expensive program that you will use only once (hopefully).

- **Support** – As with any product you buy, great support is always welcome if you encounter any problems or don’t understand something. The more personalized the support, the more applicable it is. Many people have been frustrated with call centres that read from a pre-defined list of questions but never get to the heart or fully understand your problem. How do you know if a company's support will be to the level you expect? Simple. Test it by contacting their support with a question before you buy their product.

Many file recovery tools may check only a few of the boxes above but the best, like Media Investigator will check all of the above boxes.

Conclusion

We have seen how digital data is in almost every area of our modern lives and how that data can become lost due to human error or equipment malfunction.

We have seen how that data is technically stored on digital media and some of the different software mechanisms used for recovering that data.

We have also seen how you can maximize your chances of a successful recovery by taking some simple steps and, if you should lose your data, what to look for when selecting a file recovery tool.

One of the most important things to realise is that all may not be lost and your data may still be there even after a file deletion or drive format which is useful in today’s world where many of our memories, contacts, finances, and music are stored digitally.

This article was written by Data Recovery Systems Ltd., creator of the file recovery program Media Investigator which can be downloaded at http://www.DigitalFilerecovery.com.