

## A Digital Forensics Primer

A little understanding of electronic evidence and digital forensics goes a long way. Because there are some terms of art that mean one thing to a forensic investigator and another to a layperson, it is important that you familiarize yourself with a bit of the lexicon before engaging a digital forensics firm. This primer will help with the most commonly misunderstood terms.

### Acquisition

When we talk about acquiring evidence in forensic investigations, we aren't talking about receiving it. An art dealer may say "I acquired a rare Picasso while in London" – meaning he took possession of it. When a digital investigator talks about "acquisition", they mean obtaining a forensically-sound copy of the evidence. This may be either an "image" or a "clone" – both defined below.

### Image

When we talk about an image, we are talking about a bit-by-bit copy of the source material into a file (or series of files) to be used in the investigation. The image files are not accessible without specialized software and some popular formats support compression and encryption. You may hear images talked about in "flat-file" format – where a 20GB drive produces a 20GB file, which can also be split into segments for more convenient storage. All popular e-discovery and forensics platforms can read flat files. You may also hear about "EnCase" or "E01" files. These are a compressible, encryptable format for use in Guidance Software's EnCase investigation software. Accessing an image does not modify the data it contains and no specialized hardware is needed.

### Clone

A clone is a copy of one hard drive to another. It is readable in the same way the original drive is and can be put in an enclosure and connected via USB for perusal. It is not uncommon to create both an image for the forensic investigator and a clone for the client to look through. A clone will be modified if it is not accessed through specialized hardware that prevents writes to the disk (a write-blocker).

### Hash

Not the shredded potatoes, but a mathematical function used to fingerprint a digital file or disk. The most popular are MD5, SHA-1 and SHA-256. You may hear some discussion of MD5 and SHA-1 being "broken" – but this vulnerability is mostly theoretical insofar as its application in e-discovery. The "weakest" of the three – MD5 – only has a 1 in 340 trillion trillion trillion chance of being inaccurate. By comparing hash values, we can identify matching files very quickly. We can also use it to verify that a data has remained unchanged by comparing the original hash value with the current one.

### Carving

Carving is the process by which deleted file can be recovered long after the computer's file system has forgotten about them. You may also hear this referred to as "raw recovery" sometimes. Most files have a defined structure. By searching through the media for this "file signature" it is possible to recover fragments of files or even entire files years after they were deleted.

### Unallocated Space

When you save a file to disk, the computer makes an "allocation" of space on the disk for that file. When you delete that file, the entry corresponding to it is removed from the allocation table. That space is now unallocated. Unallocated space is that area of the disk that the file system has marked as available for use. It is often possible to recover hundreds or thousands of files from unallocated space.

### Slack

The area of a disk is divided into units called clusters. Files start at the beginning of a cluster for ease of organization. If a file is not precisely the size of a cluster (a rarity to be sure) then there will be

## A Digital Forensics Primer

some space left between the end of the file and the beginning of the next cluster. This is slack. It is possible that remnants of a previous file will be readable from slack space on a disk.

### Metadata

Metadata is simply “data about data.” There are two types commonly referred to: filesystem metadata and program metadata. Filesystem metadata includes the security permissions, dates of last access, last write and creation and whether a file is hidden, compressed or archived. Program metadata is information written into a file by an application. In the case of Microsoft Word, this can include the Author, Organization, Date of Last Printing and even a journal of changes and commentary in a collaborative document.

### Wipe

Also called sanitizing, scrubbing, erasing, zeroing and many other things, wiping a drive involves overwriting every location on the drive with new information. Because each location can only contain one value at a time, overwriting renders the previous data unrecoverable. As many of my colleagues and I have been saying for quite some time, regardless of what old wives’ tales you may hear, a single-pass overwriting of data renders it permanently and irrevocably unrecoverable. You may hear of three-pass, seven-pass and even thirty-five-pass erasure. This is unnecessary overkill. One pass is sufficient to defeat all known methods of data recovery.