Teaching Case

Data Storage Forensics – What is Really Left After I Hit the Delete Button, and How Can I Actually Make Sure It's Gone?

Anthony Serapiglia
Anthony.Serapiglia@stvincent.edu
CIS Department, St. Vincent College
Latrobe, PA 15650

Abstract

The following Teaching Case is designed to expose students to three scenarios related to data stored on hard drives, techniques that could be used to retrieve deleted or corrupted data, and a method for a more thorough deletion of data from a hard drive. These issues are often overlooked in current IT curriculum and in our age of digital clutter this can be a dangerous oversight leading to potential financial loss, exposure to identity theft, and criminal liability. This case study / lab exercise can be utilized in multiple levels of a CS/IS curriculum, adjusted to meet the skill and background levels of introductory courses to specialized capstone courses in hardware or security. It provides talking points to highlight the importance of being aware of the spreading digital footprint, and provides introductory exposure to available tools and techniques for advanced data recovery.

Keywords: Computer Forensics, Data Recovery, Disaster Recovery, Identity Theft, Digital Footprint, Hard Drive, Drive Imaging

1. Introduction

Our lives today are digital. News of Big Data and the amounts of information that are collected and stored abound till the numbers simply make us numb and are inconceivable of what they really mean. One report states that 90% of all of the data in the world has been generated in the past 2 years (Dragland, 2013). To survive today it is necessary to know how to cope with, work with, and become efficient with large amounts of data. It also mandates that just as with real trash, we know how to deal with our digital trash.

The term Digital Footprint is often used in relation to a person's online presence. However, it can be extended to include much more than traces of social media posts. Just as dumpster divers can gather much information about a person through their physical trash, so too can digital evidence be gathered through the physical collection of digital trash and physical storage mediums such as hard drives, flash drives, and SIM cards. Simply assuming a corrupted drive is inaccessible or that the drive has been re-formatted so everything must be deleted is a mistake. Possibly, you have left much behind.

A normal hard drive keeps an index of where a file is placed within its storage space. This is the same regardless of operating system (OS), MAC, Windows, or LINUX. The normal deletion process

is to simply remove the pointers from this index that tells the OS where the file is. Thus the OS no longer thinks the file is there. However, the file is, normally, still there on the storage medium and will remain so until it is overwritten by another file. Even when overwritten, a file is often split into various chunks and pieces and scattered across the available space. This can lead to partial files remaining available even if some sections have been overwritten. File carving is a practice of partial recovery of files and can still reveal very valuable material even if the full file is not available. Discarded hard drives and other storage devices have become a digital dumpster diver's dream.

It is important for anyone working in the IT industry to understand the extent of their digital footprint and to manage sensitive data securely. This includes being informed on how to properly delete and destroy data, as well as how to possibly recover accidentally deleted or corrupted data. An entire industry exists of companies that can perform these tasks at a cost, sometimes a very high cost. There does exist, however, a multitude of free and readily available tools that can allow anyone to perform data recover and secure data deletion quickly and easily

Included in this paper is a procedure for a lab exercise that will walk students through the process of collecting an image of a hard drive and investigating that image for any data it may contain. The lab includes the use of several software packages that are free, and can be adapted for several operating systems or other forensic tools. While secure data collection for evidentiary use will be discussed, it is not necessary for the scope of this exercise. An activity can be included to compare various deletion methods to highlight the need for care in data disposal. These exercises have been utilized in a non-major introductory course, a course on computer architecture and operating systems, and an advanced computer security course. The exercises and exploration questions can be adjusted and presented to be applicable for students with a wide range of computer and technological experience and expertise.

2. Three Background Descriptions

The following three scenario outlines have been provided as entry points to stimulate conversation, provide a personal attachment for students to relate to similar experiences in their own lives, and highlight the idea that the amount of places data resides is vast and often not very obvious.

Personal Data Loss

There is a common phenomenon that happens today in regard to the changing mediums of our lives. It centers on family gatherings. These could be happy occasions, such as weddings or reunions, or sadder occasions such as funerals. Whatever the occasion, there is often a presence when generations of a family come together that have not seen each other for extended periods of time – that presence is "the box" or "the albums"… pictures. Pictures in black and white or sepia tone, old Polaroid's, faded prints from the 1960's, 1970's, and 1980's populate the albums and boxes. They are passed around and handled with much care as they are precious artifacts of life. A funny thing happens with those boxes and albums, though. They tend to stop. They all of a sudden drop off around the year 2000. Many people have switched from film and prints to digital cameras and phones with cameras built in. Cost, storage, and convenience have combined to change our habits related to these memories.

A staple of the nightly news, unfortunately, is coverage of a house fire. In the background will be a scene littered with fire trucks, ladders, hoses, and firemen. A reporter will find a resident who is safe from harm, but distraught over the loss. "All of our memories are gone!" they will say. All of those albums and boxes of photographs burned and lost to the fire. With the change in medium today to digital files and storage in hard drives, the same loss can occur in an instant without the fire. The most common point of failure in a computer is the hard drive. The most common drives to fail are large capacity consumer models for home use. Thousands of digital images could be lost in an instant. There are no "film at eleven" news crews to cover this tragedy, but the loss is felt by the owner nonetheless.

Questions

How do you backup your files? Do you utilize cloud services such as Dropbox, Google Drive, or Microsoft SkyDrive? What are some recovery services available that will find data one equipment that has suffered fire/water/physical damage?

Digital Footprints in Unexpected Places

On April 10, 2010 CBS Evening News aired a 5 minute investigative reporting piece that provided just another piece of evidence that our digital

footprints are large and at the same time mostly hidden from our primary sight. The piece by reporter Armen Keteyian came during the 50th anniversary of the ubiquitous steady workhorse of office machinery – the copy machine (Keteyian, 2010). In the fifty intervening years the copy machine had undergone drastic evolution from a more manual piece of machinery such as mimeograph machines, to devices that are as complex and powerful as standalone personal computers. The reality is, since the early 2000's, almost every multifunction printer/scanner/copier (MFP, multi-function printer) is a standalone computer that has a processor, an operating system, and a hard drive as the basis for the machine. Because of their status as a utility device tucked away in a separate room or closet, the MFP is rarely seen by those that use it as a computer – it is just another background object that sometimes needs maintenance, sometimes needs refilling, but falls right in line with the water cooler in how much attention it really gets.

During the course of the investigative reporting piece by Mr. Keteyian, it was shown that a grave vulnerability lies within these multifunction printers – the hard drive. Most office workers do not know that the MFP they use so often operates in such a fashion that each copy is actually a scan that is saved to this hard drive so that the printer side of the device can output multiple hard copies without re-scanning every time. Unfortunately not only do most users not know that the image file is even there, most MFP's also keep this file on the hard drive until space is filled and the need to overwrite existing files is reached.

MFP's have become very expensive pieces of office equipment. Most companies, no matter what size, do not own their own opting rather to lease these machines from office supply companies or even the manufacturer directly. Often they are rotated out after three to five years. The lifespan of these MFP's can be ten to fifteen years or more with proper maintenance and care. The aftermarket for used machines is great with used machines being in demand for their cost savings over new models. Thus it is not uncommon to see a ten to fifteen year old copier that has been in four or more different offices for different companies over time.

The vulnerability that the CBS News investigative report highlighted was that in most cases, being unaware of the existence of the hard drive, MFP's that were recycled to other locations or sold to other office product suppliers rarely, if ever, cleared the hard drive of the accumulated images of documents that had been printed from it. In the report, 4 copiers were purchased from a used equipment warehouse for an average of $300 each. Through freely available forensics software, all four offered up a treasure trove of sensitive documents. One machine was from a sex crimes division of a metropolitan police force and held documents related to criminal cases. A second was found to be from another police department narcotics division that contained documents that included details of suspects in drug raids amongst other information. The third had been used in an architectural firm and it contained structural design plans for buildings in Manhattan, one a block from Ground Zero as well as 95 pages of Human Resource documents for payroll that included social security numbers and pay stubs. The final MFP had been placed with a medical insurance company and produced over 300 pages of private medical records including prescriptions, blood test results, and a cancer diagnosis.

It was revealed during the report that although some manufactures are aware of the issue of saved images on hard drives, many have done nothing about it. Some have made available software updates, at a cost as high as an extra $500. With the life span of older machines being extended in the aftermarket to well over 10 years, this is a problem that will not go away soon.

Questions

How many place does your data live? If you sent an e-mail with a spreadsheet attachment, itemize how many different places that spreadsheet can end up. How much data is stored on your cell phone? If you lost your phone, or if it was stolen – how much of that information is sensitive? Would you have a method of remote deletion?

Evidence

Statistics provided by the 2012 Norton Cybercrime Report shows the rising scale of consumer cybercrime. The numbers are staggering, showing 1.5 million victims daily averaging 18 victims per second. The global price tag of this crime is estimated at $110 billion US dollars annually (Symantec, 2012). Computer crime, in general, is a very vague topic. There are very few set definitions and boundaries within law enforcement. Multiple descriptions exist with multiple standards organizations offering up "best practices" guidelines on evidence collection and handling. The US department of Justice categorizes computer crime in three categories:

Where the computer is the target of the crime, where the computer is used as a weapon in commission of a crime, and where the computer is used as an accessory in commission of a crime (USDOJ, 2013). Considering these broad definitions of computer crime in combination with the broad definition of what a computer can be - one estimate shows 90% of legal cases in 2008 included digital evidence of some sort (Science Daily, 2009).

Issues abound when dealing with digital evidence. While best practice recommendations help, the final decision on admissibility often lays within the decision of an individual judge. Also, once admitted, the value of that evidence can sway greatly with the presentation by an adept lawyer, as well as the perception of a jury who may or may not bring with them expectations of magic based on Hollywood portrayals of the magic of Information Technology. Research has been done on the "CSI-Effect", how the portrayal of technology in television crime shows effects potential jurors (Davis, Paullet, et al, 2010). This and other studies (Overill, 2013; Slaught, 2013; Hayes, Leavett 2013; Cole, 2013) concludes that those who watch a large amount (4 hours a week or more) of crime related shows, often exhibit un-real expectations in regard to the capabilities of technology in relation to crime scene investigation, and are less likely to answer knowledge questions on forensic technology correctly than those who watch less crime related television. Other studies have shown that law enforcement, though, has altered their practices in response to the perceived "CSI-Effect" and expectations of potential jurors in regard to amount and accuracy of collected evidence (Kopaki, 2013).

With the amount of cases involving digital evidence and the expectations of jurors to have that evidence collected and presented perfectly, the need for trained technicians who can present themselves and the evidence in a clear and understandable manner is tremendous.

Questions

Is there a certification process to verify experts in trials have been properly trained in handling digital evidence? What is the percentage of law enforcement personnel who have been trained in digital evidence collection and interpretation? What is a "hash" when imaging a hard drive? What are multiple methods that could be used to write protect a hard drive while imaging? If a hash changes, does it make any evidence inadmissible in a court of law?

3. Jobs

Analysts who work for state or federal law enforcement agencies usually earn a starting salary of between $50,000 and $75,000. Salary can increase with experience, advanced degrees, and security clearance.

Computer Forensics Investigators or Forensic Analysts are in high demand. With abundant opportunities in both public and private sectors, the job outlook is excellent. The US Department of Labor projects a growth rate of over 20% between 2010 and 2020, placing the profession in the projected top 10 percent of growth professions. (BoLS, 2013)

4. Lab Exercises

The following exercises attached (Appendix A) can be performed together, or broken into parts to correspond with ability and level of the course. There are two major tasks involved, taking an image of the hard drive and recovering files from the image. These tasks can be taken separately, or as a series. One possibility for condensing assignments would be to have an instructor take an image of a hard drive ahead of time and have students work individually analyzing the image provided to them.

5. Bibliography

Cole, Simone (2013, June). A surfeit of science: The "CSI effect" and the media appropriation of the public understanding of science. Journal of Public Understanding of Science. Retrieved from http://pus.sagepub.com/content/early/2013/04/09/0963662513481294.abstract

Davis, Gary; Paullet, Karen; Houck, Max; Swan, Tom (2010) Does the Technology Portrayed In Television Crime Shows Have an Effect On Potential Jurors? Issues in Information Systems, *Volume XI, No. 1.* Retrieved from http://iacis.org/iis/2010/154-163_LV2010_1439.pdf

Dragland, Ase (2013, May 22). Big Data, for better or worse: 90% of world's data generated over last two years. *ScienceDaily*. Retrieved from

http://www.sciencedaily.com/releases/2013/05/130522085217.htm

Hayes, Rebecca; Leavett, Lora (2013, June) Community Members' Perception of the CSI Effect. American Journal of Criminal Justice Volume 38, Issue 2, pp 216-235. Retrieved from http://link.springer.com/article/10.1007/s12103-012-9166-2

Kassner,Michael, (2010, June 14). The truth about copier hard drives: Tips for securing your data. The Tech Republic, Retrieved from http://www.techrepublic.com/blog/it-security/the-truth-about-copier-hard-drives-tips-for-securing-your-data/

Ketyean, Armen. (2010, April 19) Copy Machines, a Security Risk? CBS Evening News, Retrieved from http://www.cbsnews.com/video/watch/?id=6412572n

Kopacki, Christopher (2013, August 12). Examing the CSI Effect and the Influence of Forensic Crime Television on Future Jurors. Dissertation Defense Virginia Commonwealth University, retrieved from https://dizzyg.uls.vcu.edu/handle/10156/4465

Overill, Richard (2013). The 'inverse CSI effect': further evidence from e-crime data**.** *Int. J. Electronic Security and Digital Forensics*, 5, 81-89

Norton Security (2012, September 4). The 2012 Norton Cybercrime Report. Norton/Symantec, Retrieved from http://now-static.norton.com/now/en/pu/images/Promotions/2012/cybercrimeReport/2012_Norton_Cybercrime_Report_Master_FINAL_050912.pdf

SIMTEF (2009, January 1) Digital Evidence: Cyber Forensic Researchers Make The Call. Retrieved from http://www.sciencedaily.com/videos/2009/0104-digital_evidence.htm

Slaught, (2013) The Real CSI: A Criticism of Media's Manipulation of Forensic Science. Dissertation Defense. Retrieved from http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1142&context=comssp&sei-redir=1&referer=http%3A%2F%2Fscholar.google.com%2Fscholar%3Fq%3Dcsi%2Beffect%2Bin%2Bthe%2Bcourtroom%26btnG%3D%26hl%3Den%26as_sdt%3D0%252C39%26as_ylo%3D2013%26as_vis%3D1#search=%22csi%20effect%20courtroom%22

US Bureau of Labor Statistics (2013) Occupational Outlook Handbook, Forensic Science Technicians. Retrieved from http://www.bls.gov/ooh/life-physical-and-social-science/forensic-science-technicians.htm

USDOJ (2013) United States Department of Justice, Computer Crime and Intellectual Property Section. Retrieved from http://www.justice.gov/criminal/cybercrime/

Appendix A

The following exercises can be performed together, or broken into parts to correspond with ability and level of the course. There are two major tasks involved, taking an image of the hard drive and recovering files from the image.

Task 1: Taking an image of the hard drive.

Needed: Hard drive, power to hard drive, data connection cable, imaging software.

Hard drive: Any hard drive will do, even failed drives with bad sectors. IDE and SATA drives are most commonly available and most easily connected to a workstation.

Connection: For these exercises, we will assume that we are not completing what would be considered a legal forensics copy of the drive (details can be covered in subsequent follow up research). As such, direct connection to a workstation through internal data cables and internal power supply connectors can be accomplished. However, if such access is unavailable due to security locks or other restrictions, external USB connection cables with a separate power supply can be purchased at a current estimated cost of under $5 each.

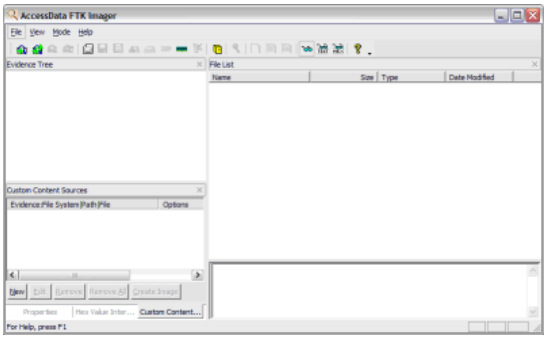
Imaging Software: There are many tools available to achieve this task. This paper will provide a step by step procedure utilizing FTK Imager from Access Data. This application is available in many other compilation packages of forensics tools, including the Forensic Tool Kit also available from Access Data. It is also available as a standalone program:
http://www.accessdata.com/support/product-downloads#Utilities

Assumption: The following procedure describes the creation of a drive image utilizing FTK Imager 3.1, a Windows 7 host machine, an external USB connection cable, and a 20GB IDE hard drive.

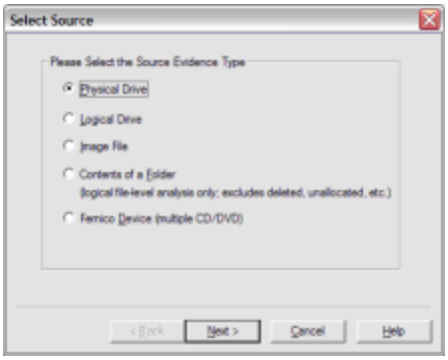
1 – Prepare the hard drive for connection. The drive may or may not have a jumper that manually sets the behavior of the drive when connected to act as a master or slave drive. Most often, removing the jumper will cause the drive to act in the mode of "cable select" and leave the device open for connection without interfering with any existing drive configurations of the host machine. This is of much greater importance if connecting the drive internally.

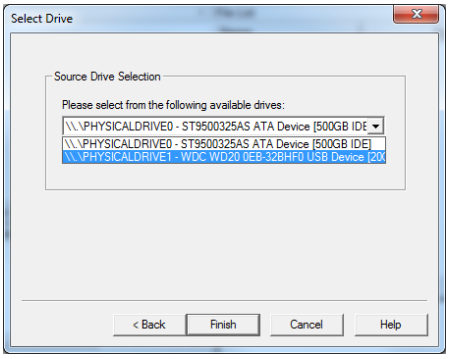
2 – Connect the drive. Connect the data cable and then the power cable. Power up the machine if it is not already on. The hard drive should be automatically found by a windows based machine.

3 – Run **FTK Imager.exe** to start the tool.

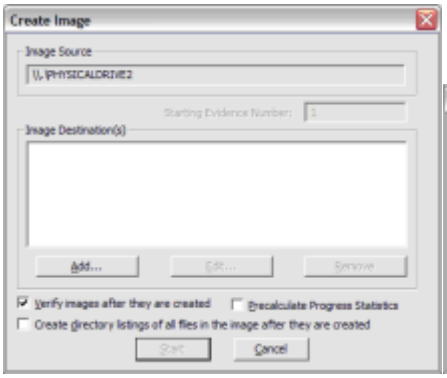


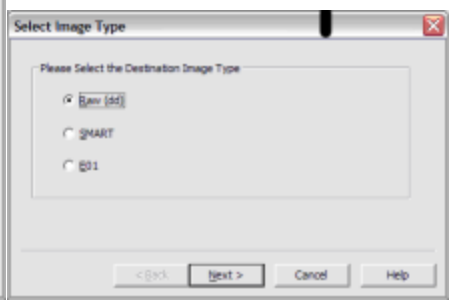
4 - From the **File** menu, select **Create a Disk Image** and choose the source of your image – **Physical Drive**. After clicking NEXT, you should see at least two options in the drop down menu, the system drive of the host machine, as well as the attached external drive. In this case, "PHYSICALDRIVE1" is also tagged as being a "USB device" for easy identification. Click **Finish.**



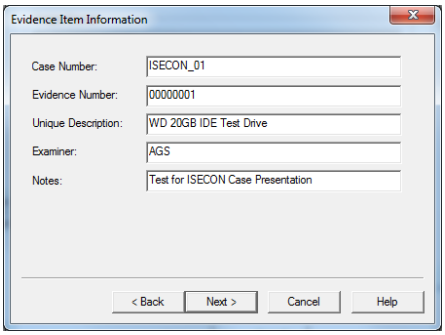


5 - To direct the image destination in the next screen, click **Add**. For greater flexibility in recovery software later, choose the **Raw (dd)** format and **Next**. Check **Verify images after they are created** so FTK Imager will calculate MD5 and SHA1 hashes of the acquired image.

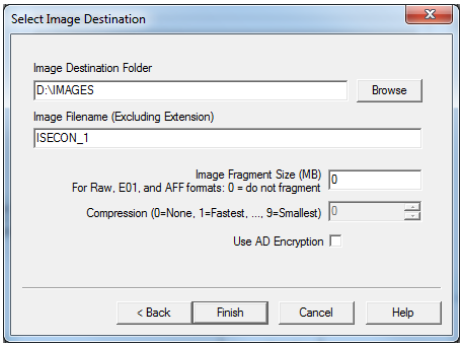




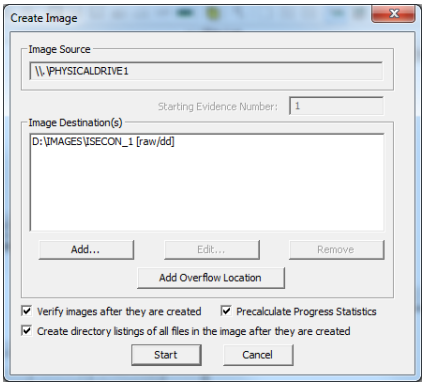
6 – Provide some information to identify the drive and your imaging session. If you select raw (dd) format, the image meta data will *not* be stored in the image file itself. A text file log will be created at the end of the process. Select **Next**.



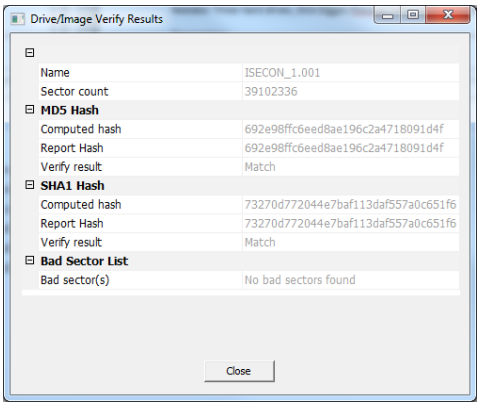
7 - Select the Image Destination folder and file name. You can also set the maximum fragment size of image split files for large capacity hard drives. For this example, enter "0" to create one image file. Click Finish to complete the wizard.



8 - Click **START** to begin the acquisition:



A progress window will appear. For this example, a 20GB IDE drive was imaged in approximately 45 minutes. Once the acquisition is complete, you can view an image summary and the drive will appear in the evidence list in the left hand side of the main FTK Imager window. You can right-click on the drive name to Verify the Image:



FTK Imager also creates a log of the acquisition process and places it in the same directory as the image, ***image-name*.txt**. The file will list the evidence information, details of the drive, check sums, and times the image acquisition started and finished:

Created By AccessData® FTK® Imager 3.1.3.2
Case Information:
Acquired using: ADI3.1.3.2
Case Number: ISECON_01
Evidence Number: 00000001
Unique description: WD 20GB IDE Test Drive
Examiner: AGS
Notes: Test for ISECON Case Presentation

Information for D:\IMAGES\ISECON_1:

Physical Evidentiary Item (Source) Information:
[Device Info]
Source Type: Physical
[Drive Geometry]
Cylinders: 2,434
Tracks per Cylinder: 255
Sectors per Track: 63
Bytes per Sector: 512
Sector Count: 39,102,336
[Physical Drive Information]
Drive Model: WDC WD20 0EB-32BHF0 USB Device
Drive Serial Number: 152D20337A0C
Drive Interface Type: USB
Removable drive: False
Source data size: 19092 MB
Sector count: 39102336

ATTENTION:
The following sector(s) on the source drive could not be read:
1091244
The contents of these sectors were replaced with zeros in the image.

[Computed Hashes]
MD5 checksum: 692e98ffc6eed8ae196c2a4718091d4f
SHA1 checksum: 73270d772044e7baf113daf557a0c651f6155602

Image Information:
Acquisition started: Mon Aug 19 11:00:23 2013
Acquisition finished: Mon Aug 19 11:45:13 2013
Segment list:
D:\IMAGES\ISECON_1.001

Questions:

1 – What steps need to be performed to ensure a sample hard drive does not have any data written to it during the image taking process?

2 – Can any drive be imaged? What steps may be taken to image a drive that may be damaged or is not recognized by Windows?

3 – What can be used to verify that a hard drive has not been tampered with, or that anything has changed after an initial image has been taken?

4 – If there is confirmation that the drive contents have changed between an initial imaging and later testing, would this automatically preclude the drive from being used as evidence? Explain.

Task 2: Recovery Comparisons

Needed: Drive image, Disk Digger/Autopsy/ ReclaiMe /other data recovery tool

http://diskdigger.org/download

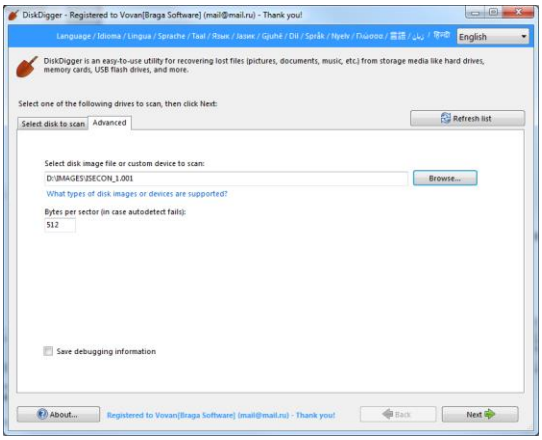
http://sourceforge.net/projects/autopsy/files/autopsy/3.0.6/

http://www.reclaime.com

http://killdisk.com/downloadfree.htm

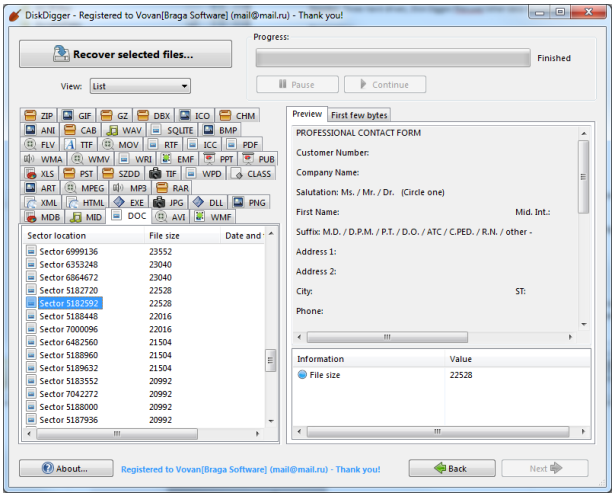
Recovering files with Disk Digger –

1 – Run DiskDigger.exe to start the tool. Click on the Advanced tab to load the drive image file. Browse for the file and select Next.



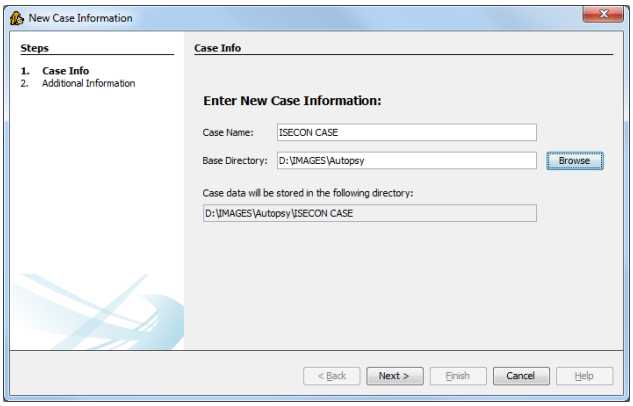
2 – For this exercise choose Dig Deeper. Continue with default selections. Selecting next will start the scan. This example scan completed in approximately 10 minutes.

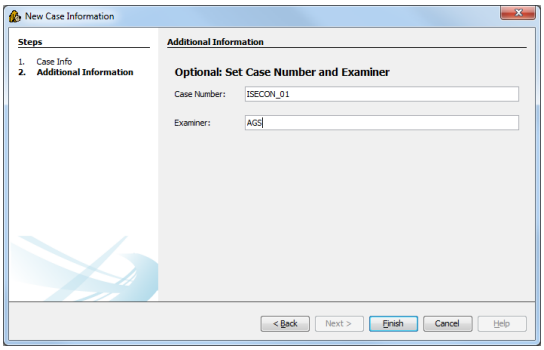
3 – During the scan and once completed, the Disk Digger interface allows for a preview of found files. Selected files can also be exported and saved to a collection directory of your choosing.



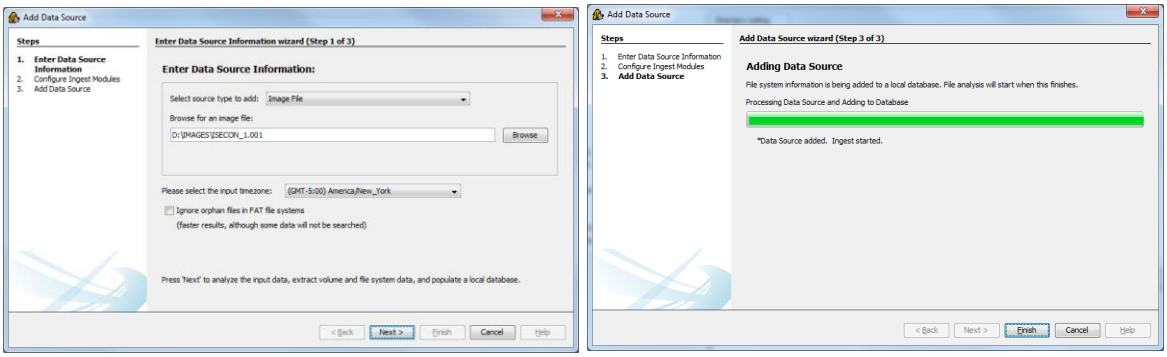
Recovering files with Autopsy

1 – Open Autopsy from the start menu and create a new case. Provide details for case name and directory to store files. You will also be asked for a case name and name of the examiner.

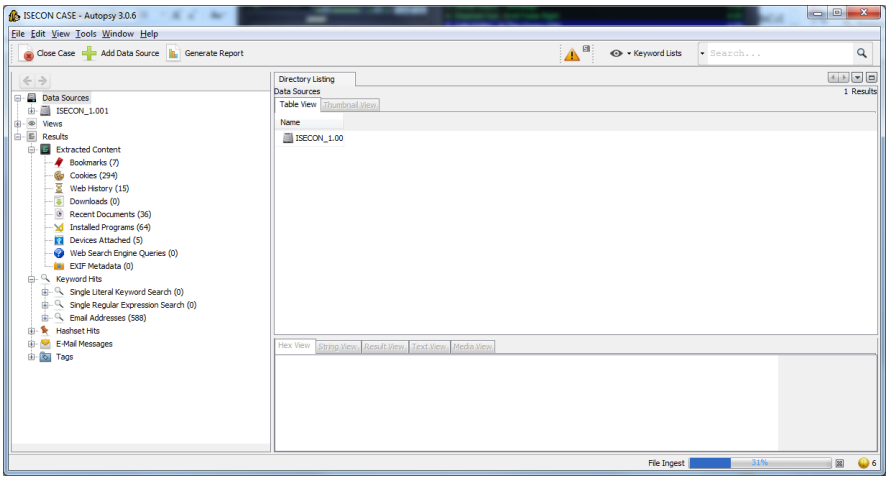




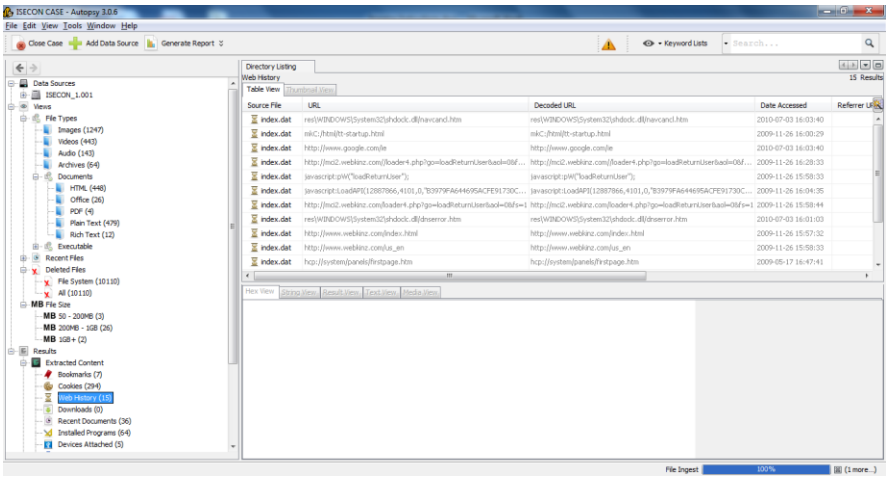
2 – Enter the path to the image file. Take default selections for ingest models. A pause will occur as the image file is added. Scanning of the image file will commence immeadiately.



3 – A progress bar can be seen in the lower right of the program. Scanning can take a considerable amount of time.



4 – The final report from Autopsy is arranged differently than Disk Digger as the purpose for the program is different. Results are arranged in more of a report fashion with categories laid out for more investigative purposes rather than file recovery

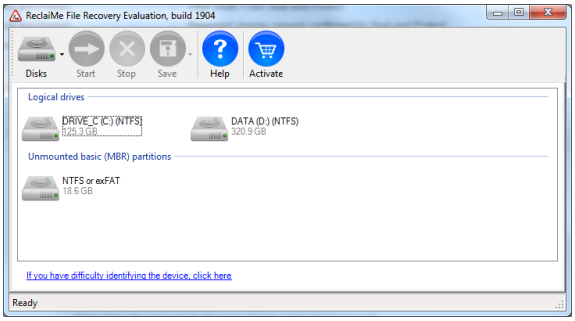


Recover with ReclaiMe:

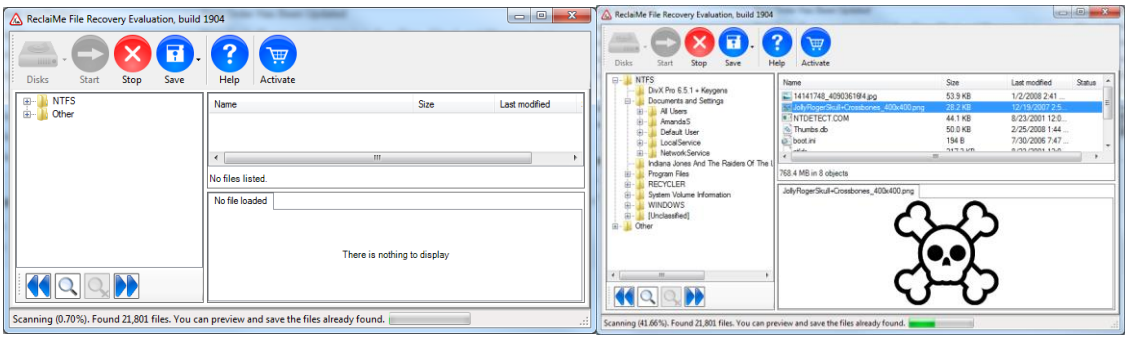
1 - Start program

2 – The program will automatically find the existing hard drives. Add your image file by clicking on the "Disks" icon in the top left tool bar. Browse for the folder the image is in. You may need to set th file type to "All Files" as ReclaiMe will work with dd images, but does not automatically find the find from FTK Imager tagged with a "001" file extension.

3 – To start the scan, double click on the newly shown "un-mounted drive" or highlight and choose "Start".



4 – A new window will appear and show a progress bar at the bottom left corner. A preview of the found files can be seen while the scan is still running.



Task 3 - Deletion Comparison:

Needed: 3 hard drives, preferably of similar size and previous use

Task: Compare recovery possibilities from three different hard drives

Setup: Prepare the three drives as A) simply removed from a computer with no deletion or formatting; B) perform a "quick format" process on the drive; and C) run the drive through a third party tool such as Active Kill Disk

Activity: Ask students to image all three drives and attempt to recover any files from the drives.

Questions:

1 – What was the approximate number of files found for each drive?

2 – Have students find and evaluate three other hard drive erasing products.

3 – Have students write up a lab report in a proper lab report format. The report can be for one, or a comparison of multiple methods of retrieval and deletion. Proper formatting, section identification, and conclusions should be evident.