Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Information Security and Communication Technology

Stian Petersen

Deepthought - A Case Study in Digital Forensic Tool Validation

Master's thesis in Master in Information Security
Supervisor: Katrin Franke, Fergus T. Toolan
June 2019

NTNU
Norwegian University of
Science and Technology

Stian Petersen

Deepthought - A Case Study in Digital Forensic Tool Validation

Master's thesis in Master in Information Security
Supervisor: Katrin Franke, Fergus T. Toolan
June 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Information Security and Communication Technology

Abstract

Digital evidence is playing an increasingly critical role in investigations. Trusting the forensic tools to retrieve and analyze files accurately, without affecting the integrity of the evidence, is imperative for the investigators to be able to use the evidence in court. In this thesis, we identify requirements for digital forensic tools, and review existing scientific methods for validating digital forensic tools. As part of this thesis we have combined ISO standards for software testing with established standards in the field of digital forensic tool testing. The result of our combined method provides a structured method for establishing test criteria in a versatile framework for testing digital forensic tools.

With increasing amounts of digital evidence in criminal cases, there is a growing need for automated digital forensic tools to process seized devices. There are many commercial tools available on the market, but with the cost of licensing commercial software, law enforcement is forced to prioritize what tools to buy. The Freetool project seeks to develop free digital forensic tools for law enforcement, and one of these tools is Deepthought.

Deepthought is a triaging tool used for performing preliminary analysis of the seized devices. By applying our test methodology, we have conducted experiments on the Deepthought software to verify if the tool meets the requirements for digital forensic tools. Our thesis concludes that Deepthought has great potential as a triaging tool, but that the current version does not meet the requirements.

As part of the thesis we have developed datasets for specific function testing of digital forensic tools. We are in the process of making these datasets available to the community for similar functionality tests.

Through our work, we have identified test assertions for digital forensic tools aimed at child exploitation cases. In future research, this work can be further extended to include more functions, and ultimately ending in a comprehensive list of test assertions to be used when testing any digital forensic tool functionality.

Sammendrag

Bevis fra digitale enheter spiller en stadig større rolle i etterforskning av kriminalsaker. Spesialverktøyene som brukes til å sikre og hente ut bevis må være nøyaktige og påvirke de beslaglagte enhetene så lite som mulig. I denne masteroppgaven identifiserer vi hvilke krav som stilles til digitale etterforskingsverktøy, samt drøfter eksisterende metodikk for å validere at slike verktøy er pålitelige. Som en del av oppgaven har vi kombinert ISO standarder for programvaretesting med etablerte metoder for validering av digitale etterforskingsverktøy. Denne metodikken gir en strukturert måte å bryte ned verktøyets funksjoner til test kriterier, i et fleksibelt rammeverk som egner seg for testing av alle digitale etterforskingsverktøy.

Med økende mengder digitale beslag i straffesaker er det et økende behov for automatiserte digitale etterforskingsverktøy for å behandle og analysere beslag. Det finnes mange kommersielle verktøy tilgjengelig på markedet, men slik programvare kan være svært kostbart noe som gjør at etterforskingsavsnittene må prioritere hvilke verktøy de skal kjøpe inn. Freetool er et prosjekt som utvikler digitale etterforskingsverktøy gratis for politienheter. Ett av verktøyene utviklet som en del av Freetool-prosjektet er Deepthought.

Deepthought er et triage verktøy som brukes til å utføre initiell analyse av beslaglagte digitale enheter. Som en del av oppgaven har vi gjennomført eksperimenter på Deepthought-programvaren ved bruk av metodikken vi beskriver. Vår avhandling konkluderer med at Deepthought har stort potensial som triage verktøy, men at dagens versjon ikke oppfyller kravene for digitale etterforskingsverktøy.

Som en del av avhandlingen har vi utviklet datasett for funksjonstesting av digitale etterforskingsverktøy. Vi er i ferd med å gjøre disse datasettene offentlig tilgjengelige slik at andre kan gjennomføre lignende funksjonstester.

Gjennom vårt arbeid har vi identifisert testkriterier for funksjoner benyttet av digitale etterforskingsverktøy rettet mot barneovergrepssaker. Disse testkriteriene kan sammenstilles og kombineres med testkriterier identifisert gjennom andre eksperimenter på digitale etterforskingsverktøy. En slik samling testkriterier kan da benyttes når lignende programvare skal testes i fremtiden.

Acknowledgements

I would like to thank my supervisors Katrin Franke and Fergus T. Toolan for their inspiration and patient guidance through my work on this master thesis. Without their steady support and advice, this thesis would not have come to fruition.

I would also like to thank my section leader at the Norwegian National Authority for Investigation and Prosecution of Economic and Environmental Crime, Thomas Walmann, for facilitating time for me to work on this thesis, as well as lending me the equipment and software needed for executing the experiments.

Finally, I would like to thank Alan Browne at the Deepthought development team for giving me access to their software, and helping me troubleshoot problems along the way.

Stian Petersen, 01.06.2019

Table of Contents

List of figures xi
List of tables xi
1. Introduction 1
1.2 Background 1
1.3 Contributions 1
1.4 Research questions 2
1.5 Limitations 2
1.6 Thesis outline 3
2 Theoretical foundation 4
2.1 Digital forensics 4
2.2 The forensic process 5
2.3 Requirements 6
2.4 Tools of the trade 9
2.5 Deepthought 10
2.6 Tests of Digital forensic tools 12
3 Method 16
3.1 Software testing methodology 16
3.2 Choice of method 19
3.3 Application of method 20
4 Experimental design 23
4.1 Hardware considerations 23
4.2 Preparation of media 23
4.3 Installation of Freetool distribution 23
4.4 Datasets 23
4.4.1 DFR-01 25
4.4.2 EDRM 25
4.4.3 Images and videos 25
4.4.4 Email 26
4.4.5 Archives 26
4.4.6 Documents 26
4.4.7 Browser history 27
4.5 Test environment 28
4.5.1 OSX environment 28

4.5.2 Linux environment 29
4.5.3 Windows environment 29
4.5.4 Analysis machine 29
4.5.5 Write-blocker 30
4.6 Reference tool and reference sets 30
4.7 Combining results 30
4.8 Deepthought settings 31
5 Experiments and results 32
5.1 Test 1 – DFR-01 Recover one deleted non-fragmented file 32
5.1.1 Scope 32
5.1.2 Assertions 32
5.1.3 Test execution 32
5.1.4 Results 32
5.2 Test 2 – EDRM Various file format extraction 34
5.2.1 Scope 34
5.2.2 Assertions 34
5.2.3 Test execution 34
5.2.4 Results 35
5.3 Test 3 – Image and video 37
5.3.1 Scope 37
5.3.2 Assertions 37
5.3.3 Test execution 38
5.3.4 Results 38
5.4 Test 4 – Email 41
5.4.1 Scope 41
5.4.2 Assertions 41
5.4.3 Test execution 41
5.4.4 Results 42
5.5 Test 5 – Archives and encrypted files 43
5.5.1 Scope 43
5.5.2 Assertions 43
5.5.3 Test execution 43
5.5.4 Results 44
5.6 Test 6 – Documents 46
5.6.1 Scope 46

5.6.2 Assertions 46
5.6.3 Test execution 46
5.6.4 Results 47
5.7 Test 7 – Browser History 48
5.7.1 Scope 48
5.7.2 Assertions 48
5.7.3 Test execution 48
5.7.4 Results 48
5.8 Test 8 – Forensic soundness 49
5.8.1 Scope 49
5.8.2 Assertions 49
5.8.3 Test execution 49
5.8.4 Results 49
6 Discussion 50
6.1 Use of method 50
6.2 Datasets and test design 50
6.3 Result comparison 51
6.4 Test results 52
6.4.1 General findings 52
6.4.2 Deleted file recovery 52
6.4.3 Images and videos 53
6.4.4 Email 53
6.4.5 Documents 54
6.4.6 Archives 54
6.4.7 Forensic soundness 54
6.4.8 Requirements 55
7 Conclusion 56
8 Future research 57
9 References 58
10 Appendices 63

List of figures

Figure 1 - Example of results from Table 5 on page 16 in the NIST report testing Autopsy 4.6.0 (40). 13
Figure 2 - Example of result presentation from (42) page 14. 14
Figure 3 - Example of listing results from (44)page 5. 15
Figure 4 - Deepthought Classification tree. 21

List of tables

Table 1 – Datasets. 24
Table 2 - MD5 checksums of DFR-01 archives. 25
Table 3 - Browser activity. 27
Table 4 - Browser distribution on operating systems. 28
Table 5 - System setup, OSX environment. 28
Table 6 - System setup, Linux environment. 29
Table 7 - System setup, Windows environment. 29
Table 8 - System setup, analysis machine. 29
Table 9 - Test 1 results. 32
Table 10 - Test 2 results. 35
Table 11 - Test 3 results. 39
Table 12 - Test 4 results. 42
Table 13 - Test 5 results. 44
Table 14 - Test 6 results. 47
Table 15 - Test 7 results. 48
Table 16 - Test 8 results. 49

1. Introduction

1.2 Background

The time where one digital forensic tool could extract and analyze all sources of evidence is over. Digital platforms are in rapid development, and the sources of evidence are ever increasing (1). Trying to catch up, the developers of digital forensic tools have to implement new features to support new devices and software at a similar rate. Because of the frequent roll-outs of updated features, the output produced by digital forensic tools should be verified often, based on scientific methods that will pass the scrutiny of the courts (2).

There exists general test methodologies developed for testing digital forensic tools (3), and some test designs for specific digital forensic tool functions (4). The methods and test cases that have been published are not designed to test all functions implemented by digital forensic tools, and do not specify how to deduce test conditions for other functions. In this thesis, we analyze existing test methodologies, and present a structured method of deducing test criteria for any functional test.

There are few public datasets suitable for testing each function of digital forensic tools (5), so in order to design appropriate tests the investigators and developers will have to create their own datasets based on the function to be tested. In this thesis, we will use public datasets, recompiled public datasets and self-constructed datasets for conducting tests.

The increase in the number of devices and amount of data seized by law enforcement (1, 6) calls for more automated processes for acquisition and analysis of seized devices. The tool we seek to test in this thesis is designed for automated triaging focused on child exploitation cases. Child exploitation is a global problem and the introduction of the internet has made it easier to find likeminded people and share illegal content (7, 8). The consequences of not detecting the illegal material on a seized device, or for the evidence to be deemed inadmissible in court could be dire. Thus, it is imperative for law enforcement to have scientifically proven methods for validating the output of the digital forensic tools they use.

The commercial tools used in digital investigations are expensive, forcing the police departments to prioritize which software they will license based on their most pressing needs and cost. The high cost of commercial tools has triggered a project for developing free software for use by law enforcement called *Freetool* (9). In this thesis, we aim to verify if Deepthought, one of the tools developed as part of this project, is ready to be used in investigations.

1.3 Contributions

We combine techniques from software testing with test methodologies used in the digital forensic tool testing community. The adaption of methods used in this thesis provides a structured method for establishing test criteria and a versatile framework for test design

for use in testing digital forensic tools. We believe our method can be used to design tests for all digital forensic tool functions.

The digital forensic tools used by law enforcement should be tested to establish if the tools are meeting the requirements. In this thesis, we identify and discuss what are the requirements for digital forensic tools, and proceed to test if Deepthought meets these requirements.

As part of our thesis we have tested the digital forensic tool Deepthought. In this process, we have established test assertions for triage tools aimed at child exploitation cases. These test assertions can be reused for all digital forensic tools with similar functionality, and is a first step toward establishing a comprehensive list of test assertions to be used for testing all digital forensic tools.

In order to test the specified functions of Deepthought, we had to recompile public datasets for various filesystems, and develop our own data sets. We are in dialog with National Institute of Standards and Technology (NIST) and Duke Law Electronic Discovery Reference Model (EDRM) to be allowed to make these datasets public to be used in similar functional tests for other tools.

1.4 Research questions

This thesis will try to answer the following main research question:

Is Deepthought reliable and ready to be used in investigations by law enforcement?

To answer this question we have to answer the following sub-questions:

- What are the requirements for digital forensic tools?
- What methods exist to test if the tool meets the requirements?
- Do the methods implemented by Deepthought meet the requirements for digital forensic tools?

1.5 Limitations

In this thesis, we had to create some datasets ourselves in order to conduct the tests needed to validate the various functions of Deepthought. Despite Deepthought supporting multiple file systems, we will only test the functionality on the *EXT4*, *HFS+*, *NTFS* and *Fat32* filesystems. The filesystems were limited to this selection because of time restrictions, and because these are some of the most common filesystems.

On the bootable image running Deepthought, we have access to the scripts orchestrating the various features of the Deepthought tool kit. We will not examine the source code of these scripts or any of the supporting tools as a part of this thesis. We chose not use the source code when designing tests because of time considerations. We believe the time and skill required to understand the inner workings of Deepthought and its interaction with the open source tools it uses, would greatly impact how many tests we would be able to conduct as a part of this thesis. Thus, the experiments in this thesis are to be considered as black box testing of the software. More details about our choice of method is detailed in Chapter 3.

1.6 Thesis outline

This section outlines each chapter of the thesis.

- Chapter 2 presents the theoretical background for this thesis. The purpose of this chapter is to introduce digital forensics as a field, and the digital forensic process. We move on to review literature in order to identify the requirements digital forensic tools should meet. Next, we describe state of the art tools in the field, and introduce the tool to be tested in this thesis. Finally, we review existing test reports of other digital forensic tools.
- Chapter 3 introduces existing methodology in software testing, and explain our choice and application of method for conducting the tests and analysis in this thesis.
- Chapter 4 details the configuration of test environment, datasets and reference tool used in the tests and how to verify the output.
- Chapter 5 presents how each test was executed and the results of each test.
- Chapter 6 discusses the test results and the test method.
- Chapter 7 provides our conclusions.
- Chapter 8 suggests future research identified through the thesis.

2 Theoretical foundation

In this chapter, we seek to establish a general understanding of how the investigators work with the seized devices, and determine requirements for digital forensic tools. We start by explaining the field of digital forensics, and the digital forensic process. Next, we will review research to identify the requirements digital forensic tools must meet, for the tool to be considered trustworthy. We then proceed to describe Deepthought, the digital forensic tool we aim to test in this thesis. Finally, we explore state of the art tools available in computer forensics, and tests of other digital forensic tools.

2.1 Digital forensics

Digital devices are now present in almost every aspect of life, with smart phones in everyone's pockets, computers at every work desk, and smart watches reminding you to walk 6000 steps each day! Digital devices are used by criminals in a wide variety of crimes, ranging from fraud, to cybercrime, to child exploitation cases, making the devices a potential gold mine of digital evidence (1, 10). Digital devices can also be of importance to an investigation without being used in the crime directly. E.g. a mobile phone belonging to a suspect can shed light on the user's location and communication before and during the time of the crime (11). By gathering data from all relevant digital devices, the investigators can create a timeline of events, providing invaluable information for the case.

The development of digital technology raised the need of forensic methods now known as digital forensics. As described by Carrier (12), the digital forensics research conference collaborated to create a definition of digital forensics (13) as:

> *"The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."*

Because a digital device can contain so much evidence, stored in various places and formats, analyzing a digital device can be seen as entering a separate crime scene, rather than just a seized item (14). Like a physical crime scene, the digital forensic investigators preserve the digital crime scene and use special tools to gather and document items of possible evidentiary value.

The skillsets needed to keep up with all areas of digital technology has led to the separation of digital forensics into specialized digital forensic fields. Some of these fields are network forensics, mobile forensics, cloud forensics, and computer forensics. Network forensics focuses on capturing and analyzing network traffic, using specialized tools (11, 13). This field of digital forensics is usually applied to investigating cyber security events, like hacking or other network attacks.

Mobile forensics is specialized in extracting data from mobile devices, like mobile phones, tablets and smart watches. The access levels to the storage of mobile phones is usually more restricted than on computers making it more difficult for the digital forensic tools to access all areas of interest (10). Smart phones can use applications like Google Drive or Apple iCloud, providing access to online storage for digital devices. Because digital

devices can use applications that store and access data online, there is often a need for cloud forensics in addition to the other fields of digital forensics in order to extract all data (10).

Cloud forensics specializes in extracting data from sources on the internet. Acquiring data stored in the cloud could provide legal challenges, because the data might be stored in a different country or jurisdiction (10). To gain access to a suspect's cloud services, law enforcement can issue a subpoena to the service provider, or use special tools combined with credentials extracted from the suspect's devices. This shows the need for the different fields of digital forensics to cooperate and share data to be able to acquire as much relevant data as possible.

In this thesis, we will focus on computer forensics, but the underlying forensic principles of the various fields are similar. As the name suggests, computer forensics focus on extracting data from computers. When using special tools to examine computers, the investigators are often able to gain access to all local storage and memory (15). This can enable the extraction of passwords from volatile memory, and recovery of deleted data. We will go into greater detail about the forensic process in the next section.

2.2 The forensic process

There have been many proposed models describing the phases of a digital forensic investigation (16). The models vary depending on the digital forensic field and the level of detail used in the models. We will describe the forensic model suggested by the U.S. Department of Justice (DOJ) (17), because we believe this model is to the point without oversimplifying the process. The DOJ model defines four phases of a digital investigation; *Assessment*, *Acquisition*, *Examination*, and *Documenting and reporting*.

In the *assessment* phase, the seized item is assessed whether it could be of evidentiary value to the investigation, and how the examiners should proceed (17).

In the acquisition phase, the examiner extracts data from the device, using tools they have verified in advance of the extraction. The goal of acquisition is to gather evidence in a forensically sound manner, using accurate tools so the evidence may be accepted in court. A bit-by-bit copy of the device, called a physical acquisition, is the gold standard when extracting the evidence for further processing (18). The output of a physical acquisition is a *forensic image*[1], often accompanied by log files describing the settings used in the acquisition, the investigator performing the extraction, and a one-way cryptographic checksum of the forensic image. The checksum can be any one-way algorithm, but the most often used is the MD5 and SHA1 algorithms. These checksums can be used at any stage of the investigation to ensure that the forensic images have not been altered, and the integrity of the files remain.

By making a forensic image, the investigators preserve the device as it was, and all analysis can be done using the forensic image (17). In some situations, the examiners will perform a logical extraction of the device. This process can extract a selected set of files, or full partitions, but will not extract all bits of the storage media in its entirety

[1] A *forensic image*, or often referred to as just an *image* in the digital forensics community. Throughout this thesis we will use the term *forensic image*, when referring to such, to avoid confusion with photos and other graphic files processed in the tests.

(17). This method of acquisition can be useful for extracting data from a computer with full disk encryption where the user is logged in, because the files would then be in a decrypted state and will be exported in this state.

In the *Examination* phase, the examiner identifies data from the device with relevance for the case (17). The identification of possible evidence might require further processing of the extracted files, an example being the discovery of encrypted archives. To aid the investigators, the forensic images are processed using tools that index and parse all files. This way the investigators can search for text and browse the filesystems for evidence.

In the *documenting and reporting* phase the examiners combine their findings to establish a timeline of facts describing the events to be used in court (18).

In cases where time is critical for the investigation, like murder or abduction cases, the investigators can conduct triage of the seized devices to gain preliminary results at an early stage (19). During triage, investigators use special tools that acquire and analyze the device, often automatically extracting and presenting artefacts of possible evidentiary value. Triaging tools can also be used in cases that are not as time sensitive, to uncover incriminating evidence which can be used when interrogating the suspect at an early stage of the investigation (20). Every phase in the digital forensic process requires specialized tools. Some of the state of the art tools will be covered in Section 2.4.

Carrier elaborates that the goal of digital investigation is to find evidence that supports or contradicts the prosecutors' theories, as well as signs of evidence tampering (21). This is important because the prosecutor's role is also to look for evidence that can exonerate the accused.

2.3 Requirements

The judicial system is the main entity for defining forensic requirements. Prior and during a case, the court must decide whether evidence is accepted or not. Scientific evidence is often presented by an expert witness to explain how the evidence was collected, and to give an interpretation of the findings.

The Daubert v. Merrell Dow Pharmaceuticals Inc. case (22) outlined three factors a court must evaluate to determine if scientific evidence is admissible in a court. These factors are *reliability* of the evidence, if the evidence *fits* the issue in the case, and the *risk* of the evidence confusing or misleading the jury. The reliability of the evidence was specified further that the courts should consider (a) if the scientific method or technique used has been tested, (b) if the method has been published or peer reviewed, (c) the error rates of the method, (d) whether there are standards governing the operation of the technique, and (e) if the method is generally accepted in the scientific community (23).

To comply with the principles described in the Daubert ruling, the tools used by law enforcement should be tested to verify that the output is accurate and the results should be shared for review in the community.

Based on the requirements set by the judicial system, law enforcement agencies develop policies and operating procedures to meet the requirements set by the judicial system.

The U.S. Department of Justice (DOJ) published in 2004 a guideline for handling and examining digital evidence (17). In this guideline, the DOJ outlines three general forensic principles they expect of a digital forensic investigation:

- *Actions taken to secure and collect digital evidence should not affect the integrity of that evidence.*
- *Persons conducting an examination of digital evidence should be trained for that purpose.*
- *Activity relating to the seizure, examination, storage, or transfer of digital evidence should be documented, preserved, and available for review.*

For comparison to a newer guideline used in Europe, we can look to the guideline published in 2012 by The association of Chief Police Officers (ACPO). ACPO was a private company funded by police authorities in England, Whales and Northern Ireland (24). In their guideline "Good practice guide for digital evidence" (20), they propose four principles of handling digital evidence:

1. *No action taken by law enforcement agencies, persons employed within those agencies or their agents should change data which may subsequently be relied upon in court.*
2. *In circumstances where a person finds it necessary to access original data, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.*
3. *An audit trail or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.*
4. *The person in charge of the investigation has overall responsibility for ensuring that the law and these principles are adhered to.*

Comparing the principles listed by ACPO to the principles listed by the DOJ, we see that many of the principles overlap. They both stress the importance of using trained personnel, documenting the process, and that the integrity of the seized device must be preserved.

Locard's exchange principle is a well-known principle of exchange in forensics. It states: "It is impossible for a criminal to act, especially considering the intensity of a crime, without leaving traces of this presence" (25). Carrier argues that this principle is also applicable to digital evidence, for the criminal as well as for the investigator (14). Some situations require changes to be made to the device to be able to acquire the data (15, 26). If a system is running, it might have to be shut down and the hard drive to connected to a write blocker for acquisition. If the investigator inserts an external hard drive to collect evidence from the running system, the operating system will track the external hard drive, and the seized device is altered. We argue that DOJ's first principle does not take this into consideration, and propose to interpret their first principle as "actions taken to secure and collect digital evidence should affect the integrity of that

evidence as little as possible". The ACPO guideline (20) does account for the investigators having to access the original data, possibly affecting the integrity of the evidence.

We argue that for a method to be accepted in the community, as required in the Daubert Standard, the method has to be forensically sound. The term forensically sound refers to the method in which the evidence was acquired, how the evidence was handled and documented, and that the process should be reliable and repeatable (27). Duke Law defines a forensically sound process as "acquiring electronic information in a manner that ensures it is "as originally discovered" and is reliable enough to be admitted into evidence." (28). If we interpret "as originally discovered" literally, this means that all data extracted by a forensic tool must be extracted in its original format in addition to data produced during processing of files. McKimmish came to a slightly different conclusion when researching what it means for evidence to be forensically sound (29). He defined the following four criteria for the forensic process:

- The *meaning* of the data must be the same as on the acquired source. An example being the format of the presented timestamps may not match the original format as long as the timestamp is the same as in the original file.
- All *errors* encountered during the forensic process must be accounted for and the impact clearly explained. Tools have to be tested prior to being used and known errors have to be announced for the process to be reliable.
- The process must be *transparent* so the process can be independently repeated and verified. Transparency is achieved by documenting every step of the process and the tools used.
- *Experience*. The digital forensic investigators must be experienced for the process to have the needed credibility.

Comparing the forensic principles from the DOJ and ACPO to the criteria proposed by McKimmish we can see that all suggest that the evidence handling should be well documented, and performed by experienced personnel.

The DOJs first forensic principle specifies that the source of the evidence should not be altered, while McKimmish states that the meaning of the data extracted by the tool should be the same as the original. These factors add to the requirements of the digital forensic tools being used. Evidence should be extracted and presented to the investigators while affecting the source as little as possible. McKimmish also argues that any errors encountered must be clearly presented to the investigators so they can assess the impact of the error. If a digital forensic tool has encountered errors during processing, affecting the output, the evidence produced might be misleading to the court. This is one of the factors to be assessed by the court in the Daubert standard, and could possibly lead to evidence being inadmissible if not properly accounted for. These requirements should be met by all digital forensic tools.

In 2009, NIST published a paper listing requirements for digital forensic tools performing "active file identification and deleted file recovery" (30). NIST lists four requirements for digital forensic tools with the capability of active file identification and deleted file recovery:

1. *The tool shall identify all deleted File System-Object entries accessible in residual metadata.*
2. *The tool shall construct a Recovered Object for each deleted File System-Object entry accessible in residual metadata.*
3. *Each Recovered Object shall include all non-allocated data blocks identified in a residual metadata entry.*
4. *Each Recovered Object shall consist only of data blocks from the Deleted Block Pool.*

The requirements NIST outlines for file recovery could be summarized as "the tool should identify and recover all files fully, without including data from other files". We believe this is a sound requirement for file recovery tools, as long as the requirement is applied to a function's test criteria, not to the tool as a whole. We will discuss this in greater detail in Chapter 3.

In the next section, we will describe some state of the art tools used in the stages of the digital forensic process.

2.4 Tools of the trade

Each field of digital forensics has specialized tools used to extract and analyze data. In this section, we will describe some of the tools used in computer forensics. We will only discuss computer forensic tools, because the tool we aim to test is within this field.

There are a multitude of commercial and open source tools specialized for duplicating, analyzing and presenting artifacts from seized devices.

Triaging tools combine the assessment phase, acquisition phase and extraction phase to gain information at an early stage of the investigation. Detego Field triage is a triaging tool designed for extracting images, videos, chat, archives, documents and allegedly over 1000 file types from computers and storage devices (31). This software is the most state of the art triaging software we have found on the commercial market, and is designed to automatically detect encrypted files, as well as previously identified illegal files based on a database of checksums.

Acquisition tools specialize in extracting data while affecting as little of the device as possible. FTK Imager is a popular free commercial acquisition tool made by AccessData. FTK imager supports acquisition of most filesystems and can generate various forensic image formats. FTK Imager can also mount forensic images as read-only hard drives (32), allowing an investigator to browse the forensic image as if it was a hard drive on the investigator's computer. FTK imager is a windows application and does not support being executed on other operating systems.

MacQuisition is an acquisition tool for Macintosh computers. MacQuisition can be run as an executable program, or the examiners can boot into a secure acquisition environment running in the memory of the seized device. By booting into a secure acquisition

environment, the hard drive of the computer can be acquired without affecting its contents and without having to disassemble the computer (33).

Analysis tools are used for processing forensic images, and presenting the data to the investigators in a manner that makes it easier to find evidence. This is often done by parsing all files in the forensic images, indexing the content for searchability, and categorizing files for easier review. One such tool is *AD Lab* made by AccessData. This tool is designed for large scale investigations, handling millions of files, and supporting evidence from mobile phones, computers and cloud sources (34). By adding evidence from multiple sources into a single analysis tool, the investigators can connect the dots and establish a timeline of events across multiple devices.

Commercial tools for digital investigations are expensive and the use often limited by the number of licenses owned by the police department. Because the commercial tools are so expensive the police departments often have to prioritize buying licenses for a few tools that meet the most pressing needs, unable to afford licenses for all specialized tools. This leads an increasing number of investigators to develop their own scripts and tools to remedy the lack of commercial tools (9). With no established channels of coordination and communication for exchanging such tools across police departments, and country borders, the investigators are often developing software with similar functionality without knowing other developers in the community might already have a tool ready (9). The Freetool project is a project aiming to fund, structure and coordinate the development of free digital forensic tools, for use by the law enforcement community (9). The project has multiple tools in development, and a website where the users can give feedback to the developers. The Freetool project started in 2013 and is coordinated by the Centre for Cybersecurity and Cybercrime Investigation in Dublin, Ireland (35).

By running open source tools for parts of the forensic process, you are limited by the number of dedicated forensic computers, rather than licenses. Parts of the acquisition process can be executed on the seized device itself, freeing up dedicated forensic computers at the police department that would otherwise have to be occupied (36). One of the digital forensic tools under development by the Freetool project is *Deepthought*. In the next section, we will examine this tool and its features.

2.5 Deepthought

Deepthought is a free digital forensic tool built on open source software for analyzing seized computers. Deepthought is built on a bootable Linux distribution called CAINE (Computer Aided INvestigative Environment) (37). CAINE is a forensics platform with scripts and preconfigured open source tools customized for digital investigations (38). The CAINE distribution has been further optimized by the Freetool project (39), which will hereon be referenced as the *Freetool distribution*.

Deepthought started as a suite of four forensic tools built on open source software developed by Adrian Shaw at the Warwickshire Police Force, UK. Deepthought was later added to the Freetool project, and the continued development of Deepthought is done by Alan Browne at An Garda Síochána, Ireland (the Irish police). The original version consisted of six scripts aimed at six different tasks; Child exploitation cases, fraud cases, asset recovery, communication recovery, keyword search and document retrieval (36). These scripts were later combined into the single tool named Deepthought. Although the

main features of Deepthought are still aimed towards child exploitation cases, the tool can be used in many other scenarios where its functions can provide useful information.

During an investigation, the Freetool distribution can be started on a seized computer, or run on a standalone machine, letting the investigators connect seized media like USB storage devices or hard drives. The Freetool distribution uses a software write-blocker to lock all connected storage media in read-only mode (38). This preserves the integrity of the system being analyzed. Deepthought can be launched from the shortcut on the Freetool distribution or from the command line.

The Deepthought user manual (37), lists the following functionality:

- Supports *NTFS*, *FAT*, *EXT*, *ISO9660*, *HFS+, UFS*, *RAW*, *SWAP*, *FAT12*, *FAT16*, *FAT32*, *EXFAT*, *EXT2*, *EXT3*, *EXT4*, *UFS1*, *UFS2*, *YAFFS2* and *APFS* filesystems.
- Extracts data from Live, deleted and unallocated space.
- Image extraction of common images formats: *jpgs*, *gifs*, *bmps* and *pngs*. And search file signatures containing "image data", "PC bitmap" and "MS Windows icon".
- Movie extraction.
- Extraction of *Microsoft Office*, *Libreoffice* and *PDF* documents.
- Image extraction from *PDFs*.
- Web analysis – Extract web artefacts from *Android webview*, *Google Chrome*, *Firefox*, *Internet explorer*, *Safari*.
- Chat analysis supporting *Google Hangouts*, *Tango*, *Viber*, *Skype*, *Apple imessage*, *Xchat*, *Kik ios*, *Twitter*, Android call history and messages, *Facebook Messenger*, *Whatsapp* (decrypted *sqlite* only), *Gigatribe* and more.
- Email extraction of common email formats: *OST*, *PST*, *DBX*, *MSG*, *EML* and *MBOX* formats.
- Extract encrypted files larger than 500 Mb.
- Image and movie extraction from common archive formats: zip, rar, 7z etc.

When running Deepthought, the investigator is prompted to register information about case type, case number, exhibit number, investigator name and suspect name. This information is later listed in the report. The user must select the device or forensic image to be processed, and where the processed date should be stored. The investigator also has the option of adding a hash database containing checksums of known illegal content.

Next, the investigator can select the level of processing. This input field has two options, *Live/Deleted files only*, and *Live/Deleted/Unallocated*. The user manual (37) does not detail the how the evidence is handled differently, but our assumption is that the latter option includes searching for file signatures to recover files.

The investigator is provided with a list of parsing options for the investigator to decide what features are important for the particular case. The options are:

- Image extraction.
- Movie extraction.
- PDF Image extraction.
- Document extraction.
- Registry.
- Web analysis.
- Chats.
- EXIF.
- Email extraction.
- Encrypted files.

- Archive files.
- Create C4P Package

The investigator also has the option of adding a keyword list. These words will then be searched for during the processing.

Behind the scenes, Deepthought is orchestrating multiple open source tools and Linux commands to analyze the seized device. After processing the images and videos, Deepthought opens a browser where the investigator can review and tag the images and videos according to the nature of their content.

After Deepthought has processed the device it will produce a web report listing the most important findings in separate tabs. This includes information about the device, illegal images and videos, browser history and more.

In 2014, Toolan et al. (36) published a paper with initial validation of some of Deepthought's functions. At the time, Deepthought was still a suite of multiple scripts specialized to solve different cases. Since then, Deepthought has been developed further and combined into a single script with added functionality. Toolan et al. (36) outlined 7 initial tests:

1. Device detection.
2. Partition detection.
3. Evidence integrity.
4. File identification.
5. Correct file processing.
6. Recovery of deleted files.
7. File carving.

The first three tests were completed, and passed the requirements. The remaining four were not tested in the paper.

After five years of development, we will apply scientific methods of testing to verify a selection of Deepthought's functions. In the next section, we will examine how other digital forensic tools have been tested in the past.

2.6 Tests of Digital forensic tools

In our research, we had trouble finding reports detailing the tests and results of digital forensic tool testing. To gain insight in how digital forensic tools have been tested in the past, we will review some of the few of the publicly available test reports. Almost all publicly available reports have been performed by The National Institute of Standards and Technology (NIST). For this reason, to review tests executed by other organizations than NIST, one of the tests we review is quite old.

NIST Software quality group is tasked with developing tools, methods and models for testing the reliability of software. In November 2018, NIST tested the digital forensic tool Autopsy 4.6.0 (40). The particular function to be tested was the *string search* functionality. The tests conducted by NIST are part of the federated tests produced by the NIST Computer Forensic Tool Test program (CFTT). In this program, NIST provides the community access to datasets and testing frameworks for testing a range of functions used by digital forensic tools. Some of the currently available test frameworks

are for disk imaging tools, mobile device tools, string search tools, and hardware write blocker tools with more tests cases to come (4).

The dataset used for testing Autopsy is publicly available from the NIST website (41), and contains text strings of different encoding[2] and language. The strings are stored in filenames, file content and metadata of active and deleted files. The files are stored in the most common filesystems used for Windows, Linux and OSX, for testing the function's support of those filesystems. The dataset comes with a list of expected results, detailing each string, encoding, and at what sectors of the forensic image the strings can be found. The results in the test is in a table, listing the case name, expected string, and expected hits, actual hits, actual misses from active and deleted files. An example of the listed results can be seen in Figure 1.

Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01		4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0

Figure 1 - Example of results from Table 5 on page 16 in the NIST report testing Autopsy 4.6.0 (40).

In the result summary of the report, NIST explains that the datasets and test cases were generated from "frequently encountered aspects of searching for text". This indicates that the method used to create the test case is based on the test designer's experience, making it difficult to extract any useful test design methodology from this report. The dataset used in the tests comes with a list of expected results. We believe this is a sound approach for validating the output generated by the tool, and is a concept we could utilize for verifying Deepthought. The tests involved using the most common filesystems. By using more than one filesystem in the datasets, the tests will also verify that the tool behaves in the same way for all the selected filesystems.

The method of listing the results used in the report is an efficient way of reporting findings when there are many test criteria, and a similar approach might be suitable for our experiments also.

In 2014 NIST CFTT tested The Sleuth Kit (TSK) 3.2.2 (42). In this report, they tested deleted file recovery and active file listing. The test requirements are predefined in the document "Active File Identification & Deleted File Recovery Tool Specification Version 1.1" published on the NIST website and covered in Section 2.3 (43). The datasets used in this report are the *Deleted file recovery* datasets (DFR), containing files with "common file deletion scenarios" (42). Each of the test cases were repeated at least four times using different filesystems to determine how the software behaves. The forensic images contain partitions of different types of each file system. An example being the FAT forensic image contained three partitions formatted as FAT12, FAT16 and FAT32. The datasets come with a document detailing the layout of each forensic image.

The results are presented in multiple tables listing different information about the results. One of the tables covering *recovered file content analysis,* lists information about each recovered file. An example of how the results are presented can be seen in Figure 2.

[2] Text encoding defines how a computer should interpret the code for presenting the text.

Recovered File Content Analysis										
Case	**Content**	**Name**	**Size**	**First**	**Blocks**	**Tail**	**Src**	**Shift**	**Seq**	**Other**
dfr-ntfs-07	Grumium.txt	Grumium.txt	512	1	0	0	1	0	0	0
	Furud.txt	Duhr.TXT	4,096	1	7	0	1	0	0	0
	Furud.txt	Furud.txt	7,680	1	14	0	1	0	0	0

Figure 2 - Example of result presentation from (42) page 14.

The table in Figure 2 lists a test case identifier (case), source of recovered content (Content), name assigned by the tool (Name), size of content (Size), if the first block of the file is recovered (First), the number of non-initial blocks is included (Blocks), partial sector blocks included (Tail), number of source files (Src), number of times there is a shift in source file contributing the content of the file (Shift), the number of times there is a break in the secquence of blocks (Seq), and number of unidentified blocks in the content (Other). This example shows the partial recovery of a text file given the name *Duhr.TXT* originating from the original file *Furud.txt*.

In the report, NIST is using the blocks of saved data together with file size as an indicator of recovered content. This approach gives the readers an understanding of how much of the file content was recovered. We argue that this information would be more useful if the table also listed the expected results, as was the case in the report testing Autopsy (40). While the approach of using number of blocks and file size does indicate how much of each file was recovered, we believe this method would be inconvenient for use in our experiments. We believe the scope of our experiments should be to see if the tool recovers the selected files correctly. Some file categories, like images and videos, could possibly still display interesting information if partially recovered. If there are partial recoveries in our experiments, we argue that is a sufficient finding in and of itself, without having to specify to what degree the file is partially recovered. We will however manually review the images and videos if there are any partial recoveries of these file types.

The report also examines the timestamps of the recovered files. These are presented to the viewer using two tables, one with the expected results, and one with the actual results. The report lists the results and make general comments about the findings. The datasets in this experiment are designed for listing active- and recovering deleted files. Listing active and deleted files are arguably the main functionality of a triaging tool like Deepthought. One or more of the DFR datasets could be used in our experiments to test this functionality.

In 2006, Marshall Information Security and Digital Evidence (MISDE) published a report testing hashing, forensic imaging, restoring and wiping functions using the digital forensic tool EnCase version 5.05d (44). In preparing for this test, images, videos, audio files, documents and zip archives were transferred to a USB flash drive and used as input data. All files were of different formats, and described as using a known dataset. The report does not indicate that the content was used for anything other than representing data. In the test, they used a write blocker to ensure the integrity of the flash drive for the tests not involving writing to the USB flash drive. In this report MISDE used the same tool they were testing for generating the MD5 checksums used as reference for validating the output of the tool. We believe by using the same software to generate the expected data and the result data, would only verify that the functions are able to produce output,

and that the tool is consistent. If the algorithms implemented are consistently erroneous the MD5 checksums could still match.

The results are listed as tables. One table comparing MD5 checksums, and one table listing the test criteria, and a pass or fail status of the tests as shown in Figure 3. We find using the pass/fail result next to the test criteria makes the list comprehensive, and easy to get an overview of the results.

Test Number	Environment:	Actions:	Assigned Reqt's	Expected Results:	Results:
01-01	Source Media; Tableau T8 Forensic USB Bridge; EnCase Forensic v.5.05d	MD5 hash calculation performed on source drive	1	MD5 Hash calculation produced.	Pass

Figure 3 - Example of listing results from (44)page 5.

During our work on this thesis, we have been unable to find any publicly available software tests on digital forensic tools for parsing the content of files, such as browser history listing and previewing emails. Despite few available test reports, there has been some research on possible methodology on how to structure and conduct tests of digital forensic tools. In the next chapter we will go into greater detail reviewing the available test methodology.

3 Method

In this chapter, we will examine existing methods used for testing digital forensic tools and software testing in general. We will move on to discuss our choice and adaption of the existing methods for use in our experiments.

3.1 Software testing methodology

There are multiple standards governing software testing and requirements for the labs performing tests. The international Organization of Standardization (ISO) is an independent international organization developing standards covering "almost every industry" (45). One of the standards published by ISO is the ISO/IEC/IEE 29119, which describes the leading standards for software testing methods and techniques. ISO 29119 is divided into 5 sections:

- ISO 29119-1 Concepts and definitions
- ISO 29119-2 Test process
- ISO 29119-3 Test documentation
- ISO 29119-4 Test techniques
- ISO 29119-5 Keyword-driven testing

Because it is near impossible to perform exhaustive testing of software due to time and cost considerations, test cases are usually prioritized based on the importance of the function. This is the foundation of *risk-based testing* as outlined in ISO29119-1, and serves as an underlying principle in most software tests (46). Because of the importance of prioritizing what functions to test, we regard risk-based testing as a superset of the other tests described below.

Requirements-based testing is designed to make sure all software requirements are included and working as intended. The requirements are identified and used when designing tests to validate the software functions. In requirements-based testing, test cases are usually written in advance of the test execution, this practice is also known as scripted testing. The advantages of scripted testing is that the tests are easily repeatable making them well suited for verification and validation through the software lifecycle (46).

Requirements-based testing can be further separated into specification-based (black box), structure based (white box) testing and experience-based testing (47). In structure-based testing, knowledge of the source-code is used in designing the tests. Experience-based testing requires domain knowledge, knowledge of the system being tested, or experience with the previous testing (46). An example of an experience-based testing method is *error guessing*, obviously requiring advanced knowledge of the software.

Specification-based testing treats the test object as a black box, observing the output by controlling the input (47). The most relevant specification-based testing techniques for this thesis are *classification tree method*, *all combinations testing*, and *scenario testing*.

The classification tree method divides the test input data into a classification tree making it easier to identify test conditions. This is a structured way of deriving test conditions by decomposing the tool into functions, and the functions into input classes. The input classes can be further decomposed into sub-classes if needed. This way of decomposing

the input should result in non-overlapping classes within each function, where each leaf node is a test condition (47).

All combinations testing is a subset of the *combinatorial test design techniques*, and is intended to test all possible combinations of test conditions related to a function. By combining all test conditions, the tests required to reach full test coverage are greatly reduced (47).

In 2009, Guo et al. mapped the functions of a ditial forensic tool to be tested as a tree structure and decomposed each function into test criteria (48). This is the only example we have been able to find that is using this method to break down a digital forensic tool function into test conditions. This implemented test method is arguably the same as a classification tree method, although not referred to as such.

Pröll et al. argues that using a standard-conform test process, like the methods detailed in ISO 29119, could be a hindrance for the creative minds designing the tests. Using a standard-conformance test would however be desirable for customers needing the software to pass certain predefined criteria for the software to be accepted (49). We argue that testing of digital forensic tools is a field where such predefined criteria are important due to the Daubert guidelines.

Reviewing for research, we found few test designs readily available for testing all functions of digital forensic tools. Most of the papers published where digital forensic tools are tested are using general test methods without deconstructing the requirements in the same detail as suggested by ISO29119-4.

NIST has multiple projects but the ones relevant for this thesis are the Computer Forensic Tool Testing project (CFTT) (50), and the Computer Forensics Reference Data Sets project (CFReDS) (41).

NIST CFTT has developed a few specific tests for digital forensic functions in their federated testing project. These tests provide a framework for testing disk imaging tools, mobile device tools, and hardware write blocker tools (4). NIST has also made a general guideline for testing forensic tools (3) based on ISO 17025 - *General Requirements for the Competence of Testing and Calibration Laboratories.* The NIST methodology outlines 7 steps for testing computer forensic tools:

1. *Establish categories of forensic requirements;*
2. *Identify requirements for a specific category;*
3. *Develop test assertions based on requirements;*
4. *Develop test code for assertions;*
5. *Identify relevant test cases;*
6. *Develop testing procedures and method;*
7. *Report test results.*

Flandring et al. has outlined and summarized multiple methodologies for evaluating and validating digital forensic tools (51). In their work, they suggest using Becket and Slay's methodology (52) for defining requirements for digital forensic tools, and using Wildson's method of black box testing (53) for implementing the tests. Becket and Slay suggested mapping the functions and their specifications of the tool to be tested, and developing

known datasets to test the tool (52). Wildson et al. (53) proposes a test framework for black box testing consisting of 6 steps:

1. *Acquirement of software*
2. *Identification of software functionalities*
3. *Development of test cases & reference sets*
4. *Development of result acceptance spectrum*
5. *Execute tests & evaluate results*
6. *Evaluation results release*

The NIST general method is similar to the method proposed by Wildson et al. (53) with focus on the tool's individual functions. However, neither the NIST general method for testing digital forensic tools or the framework outlined by Wildson et al. covers how to identify the functions or requirements for the functions.

The Scientific Working Group on Digital Evidence (SWDGE) is a confederation of law enforcement agencies, academic- and commercial organizations tasked with generating methods and standards for handling digital evidence (54). SWDGE has released a guideline of the minimum requirements for testing digital forensic tools (55). In this guideline, they define tool categories, list what types of tests should be done in each category, and how often. For some tool categories SWDGE divides the category into a subset of functions related to that category. The most relevant category for testing Deepthought is called *forensic analysis tools*. Subsets of the forensic analysis tools category are *search tools* and *recovery tools*. This guideline however is not an exhaustive list of the various functions of analysis tools, and does not detail how to test parsing tools, for instance extracting data from databases or archives.

Guo et al. (48) state that verification and validation done by the commercial digital forensic tool vendors can be categorized as "tool oriented", while the approach suggested by NIST CFTT and SWGDE is function oriented. The difference between these categories being testing the tool as a whole, versus testing a single feature of the tool. Further, Guo et al. argues that a tool oriented validation and verification approach would invalidate the tool if one of its multiple features fail to meet the requirements, regardless of the proficiency of the other features (48).

NIST CFReDS distributes multiple datasets for testing digital forensic tools (41). Public datasets like these (56, 57) often come with a layout description detailing the content of the dataset. Becket et al. refers to a documented data set as a reference set, and underscores the importance of using a dataset with known content for testing. The reference set is used to define what results to expect from the tool being tested (58).

The SWDGE guideline suggests three methods of comparing data from the tests (55):

- Testing with a known dataset where the function is verified by comparing the input data to the output data.
- Comparison testing, where the output is compared to the output produced by a different tool.
- Empirical testing, where the input dataset is not necessarily known, but output is verified if is as expected.

3.2 Choice of method

Despite having access to the scripts running behind the scenes in Deepthought, we have chosen not to use the white box testing techniques. Deepthought is orchestrating various other open source digital forensic tools, and it would, in our opinion, be too time-consuming to understand the underlying code, and analyze the dataflow between the tools for this approach to pay off with the limited time available.

The techniques defined in ISO29119-4 are useful for deconstructing the requirements into testable conditions, yet little of the research we have reviewed has utilized any of these specific techniques. We believe this is because the test requirements might seem intuitive for the test designers.

Scenario-based testing could have been an option for our tests. Using this method would be a useful way to deconstruct the requirements based on an assumed pattern of use. We chose not to use this method because we think it will be easier to reuse the datasets for testing other digital forensic tools if we create datasets targeting as few functions as necessary for each test.

Using the publicly available dataset *EDRM File Format Data Set*, many of the functions to be tested in this thesis will be covered at the same time. This test could be viewed as an *all combinations method* on a functional level. We argue that for this to be viewed as an all combinations method, all possible combinations of test criteria must be included for the functions being tested. Hence the EDRM dataset is not fully suitable to serve as an all combinations dataset. We still choose to use this data set, opening the possibility of comparing the results produced by Deepthought to results from other digital forensic tools processing the same public data set.

To comply with ISO29119-4 we choose the *classification tree method* to decompose the functions into the supported formats listed in the user manual (37). These supported functions will then be used as requirements for the tests and corresponding files will be included as an all combinations test for the selected function.

We will use the NIST general method as an outline for executing each test. This method was chosen because it fits our choice of defining function requirements, and provides a template for developing a test plan for each function.

To compare results, we chose to use *testing with a known dataset* in combination with *comparison testing* as suggested by SWDGE. The output produced by Deepthought, the reference tools and reference set will be compared in Microsoft Excel. We opted for this solution because the output produced by Deepthought and the reference tools can easily be normalized into spreadsheets for semi-automatic comparison of data. By gathering the results from Deepthought and the reference tools in a single Excel document per test, we wish to achieve greater transparency when comparing the data, making it easier to review for the developers and third parties.

3.3 Application of method

Deepthought is aimed at triaging computers involved in child exploitation cases. The functionality we prioritize for testing should reflect the most important artifacts for finding evidence in such cases. Image and movie extraction are obviously important in child exploitation cases. The ability to locate and identify images and videos showing illegal content is an important function for tools in this category (8). The ability to list files is a basic functionality needed in all digital forensic investigations, and should be one of the functions to be tested.

The internet watch foundation (IWF) removed 105,047 web pages showing sexual abuse of children in 2018 (59). This shows one of the distribution channels used to spread this kind of illegal content. Analyzing the browser history of a suspect's machine is important in a child exploitation case to shed light on this kind of activity, so this is one of Deepthought's functions to be tested (8). To discover a suspect's contacts and communication, the ability to review email stored on the device is of interest (8). This functionality should be tested as part of this thesis. A proficient way of hiding and protecting illegal content is adding the files to an encrypted archive. The presence of large encrypted archives could be an indicator of illegal content on a suspect's machine, and should be further investigated. Deepthought supports locating and extracting encrypted archives larger then 500MB. We regard this as an interesting function to include in our experiments.

Based on the discussed functionality we have selected the following functions to test in our experiments:

- File listing
- Image extraction
- Movie extraction
- Image extraction from archives
- Document extraction
- Image extraction from documents
- Email extraction
- Encrypted file extraction

These functions were selected because we believe these are important functions for a digital forensic tool aimed at investigating child exploitation cases. Early identification and extraction of these file types could possibly have a great impact on the early stages of an investigation, as well as help identify if the device is a priority for further analysis.

We start by breaking down the selected functions to be tested into test conditions using the classification tree method. The classification tree can be viewed in Figure 4.

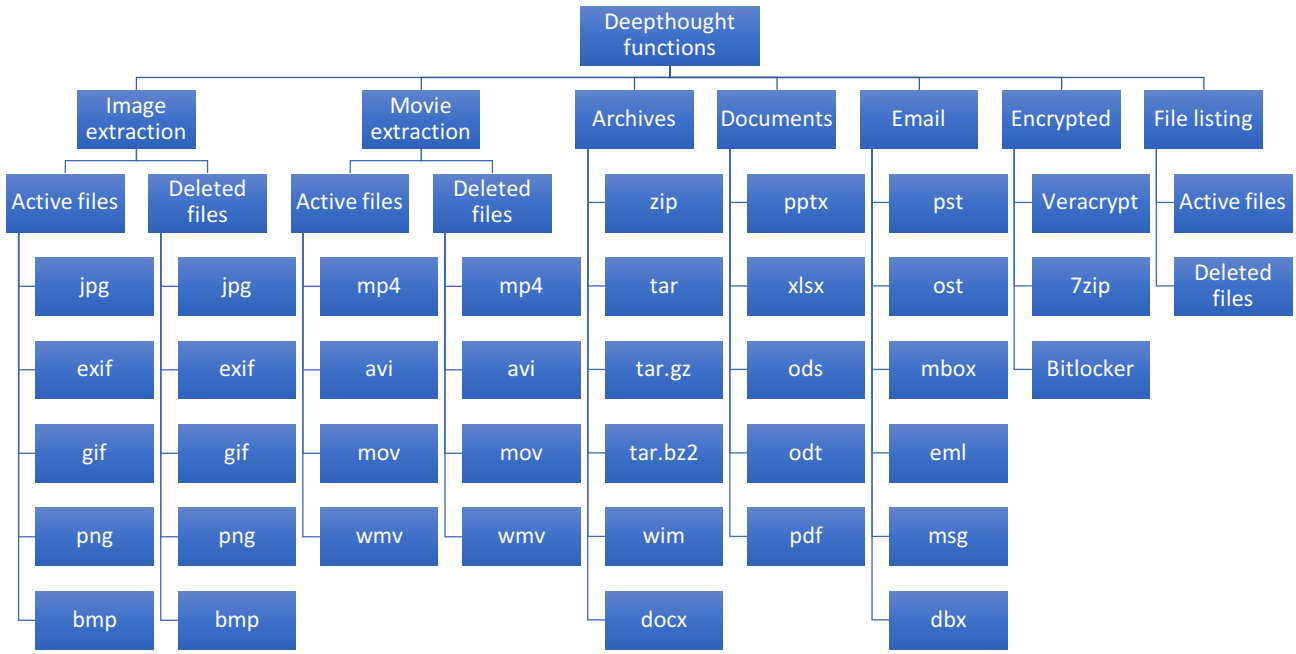


Figure 4 - Deepthought Classification tree.

We will use all combinations testing to make sure each of the test criteria (the leaf nodes of the classification tree) are covered at least once in the tests. This classification method could be used to list all possible file types within each category. We chose to only list the expected supported files based on information in the Deepthought user manual (37), because we wish to use this classification tree to define the test conditions for each function.

We will follow the approach outlined by the NIST general test methodology for computer forensic tools (3) when designing our tests:

1. *Establish categories of forensic requirements,*
2. *Identify requirements for a specific category,*
3. *Develop test assertions based on requirements,*
4. *Develop test code for assertions,*
5. *Identify relevant test cases,*
6. *Develop testing procedures and method,*
7. *Report test results.*

Steps 1 to 3 are covered using the classification tree method. In steps 4 and 5, we will create all combinations datasets, testing all test conditions for each test at the same time. Step 6 will consist of processing the datasets and comparing the results to a reference set and output produced by other tools, as suggested by SWDGE. General findings in Step 7 will be covered for each test in Chapter 5. While the Excel spreadsheets containing the full results can be found in Appendix B1-B30.

Based on the requirements in the Daubert standard and the requirements for a forensically sound process discussed in Section 2.3, digital forensic tools must meet the following requirements:

- The tool should not affect the device being investigated.
- The data presented by the tool should have the same meaning as the original data.
- Potential errors encountered during the processing should be clearly stated by the tool.

By combining the classification tree and the forensic requirements we can derive the following general test assertions for the supported files:

1. Processing a device with Deepthought should not affect the original device.
2. All deleted files should be recovered accurately.
3. All images and videos should be extracted.
4. All metadata should be listed accurately.
5. All documents should be extracted.
6. All images should be extracted from all non-encrypted archive files.
7. All encrypted archives over 500mb should be extracted as the originals.
8. All emails should be extracted as the originals.
9. All web history should be listed accurately.

These test assertions will be further deconstructed to include the leaf nodes from the classification tree in the tests. This is done to create a comprehensive list of assertions making it easier to ensure full test coverage.

In the next chapter, we will detail how we designed the datasets, and discuss considerations regarding the experiments and test design.

4 Experimental design

In this chapter, we will list the hardware and software used in the experiments. We will also detail how each dataset were created.

4.1 Hardware considerations

Most new personal computers today are configured with a solid-state drive storage unit for running the operating system. The solid-state drives (SSD) in general, have greater read and write speeds than spinning hard drives making them good test objects with regards to speed. However, the SSD's garbage collection feature can move and delete parts of the hard drive without any user interaction, resulting in different cryptographic hash-sums without any actual changes to the files on the system made by the user or the forensic tool (60). For this reason, we will not use solid state drives for our experiments in this paper. USB flash drives usually do not have controllers with garbage collection, and the variety of storage capacities make them suitable for tests with small datasets. In our experiments, we will use 8GB Kingston DataTraveler G4 USB flash drives. The Kingston DataTraveler G4 series does not have garbage collection, making it apt for our experiments (61).

4.2 Preparation of media

In order to remove any residual data before and between the tests conducted in this thesis, we wiped the media before setting up the device for each test. This process was done using the standard Windows, Linux and OSX format features and enabling full format including overwrites of the data.

4.3 Installation of Freetool distribution

The Freetool distribution ISO with MD5 checksum 269200fa157aa54bb0224f686876a6ac was downloaded from the Freetool website (39) and opened in a virtual environment using the software Virtualbox version 6.0. The Freetool distribution was then installed on a SanDisk Extreme PRO 128GB USB flash drive, following the instructions in the user manual (37). During installation, there was an error saying it was unable to install the bootloader correctly. After conferring with the developer of Deepthought, we used the built-in bootloader repair function, which succeeded in creating the bootloader.

After fixing the issue, the Freetool distribution booted as expected from the USB.

4.4 Datasets

Finding public datasets of a manageable size covering the functions we wish to test has proven difficult. In 2017, Grajeida et al. did extensive research covering 715 peer-reviewed research articles to document public datasets (5). In their research, they identified 102 datasets containing various categories of files and scenarios. Going through their findings, we found datasets we could have used to test some of the functions we selected in Chapter 3. We chose not to use these public datasets, because the datasets were either too large, containing thousands of files, or did not have all file

combinations needed for our experiments to be an all combinations test. We decided to use some public datasets, including the DFR-01 dataset from NIST (56), and image and video files published by NIST(62). As previously mentioned, we also chose to use the EDRM File Format Data Set (57) to test multiple functions at the same time.

The tests conducted in this thesis will use 7 datasets where each dataset will be transferred to at least 4 file systems[3]. Test 8 does not have a dataset in the same sense as the other tests, but is rather using the checksums of the filesystems produced in test 2, 3, 4, 5 and 7[4].

Forensic Image	Usage	Source
DFR-01	Test 1 - List one deleted non-fragmented file per file system and list active files.	NIST CFReDS[5]
EDRM_Data-Set_File-Formats_1-0	Test 2 – Listing and extraction of various file formats. Active files only.	EDRM[6] Recompiled
Images and videos	Test 3 – Recover active and deleted images and video and list EXIF data.	NIST CFReDS[7] Recompiled
Email	Test 4 – Recover email archives. Active files only.	Self-generated
Archives	Test 5 - Archives and encrypted files extraction. Active files only.	Self-generated
Documents	Test 6 – Extract documents and images from documents. Active files only.	Self-generated
Browser history	Test 7 - Parse browser history. Active files only.	Self-generated
Checksums from previous tests	Test 8 – Forensic soundness.	Self-generated

Table 1 – Datasets.

To test the desired features of Deepthought, we had to generate datasets for each function test. The process of generating each of these datasets are detailed in the following sections.

[3] The DFR-01 dataset has exFAT in addition to the HFS, NTFS, EXT and FAT filesystems.

[4] The MD5 checksums from the datasets in tests 1 and 6 are not included in test 8. The reasons for this is commented in Section 5.8.

[5] https://www.cfreds.nist.gov/dfr-test-images.html

[6] https://www.edrm.net/resources/data-sets/edrm-file-format-data-set/

[7] https://www.cfreds.nist.gov/FileCarvingTestReport/video-src.zip and https://www.cfreds.nist.gov/FileCarvingTestReport/graphic-src.zip

4.4.1 DFR-01

The DFR-01 datasets were downloaded from the NIST website (56) as individual compressed archives. The MD5 checksums of each archive is listed in Table 5.

Filename	MD5
dfr-01-ntfs.dd.bz2	fb79f0234a5920e4dabcf981e98ad3f9
dfr-01-fat.dd.bz2	1fd3f531560b48f61ae0657cadf2dcc5
dfr-01-ext.dd.bz2	1640246379c826ef93cf7510764ad101
dfr-01-osx.dd.bz2	7720ef8168c379ae7225f02fada5db6
dfr-01-xfat.dd.bz2	40d03595027d495be074826d3a6397c7

Table 2 - MD5 checksums of DFR-01 archives.

The forensic images were unpacked and transferred to the Freetool distribution for processing. NIST has created an image layout describing the files on each of the forensic images and how the forensic images were made (63). This document does not contain checksums of the files, so to the reference set will not contain checksums. In our reference set for the DFR-01 datasets, we have added the filenames found in each forensic image as detailed in the document describing the forensic image layout provided by NIST (63).

This reference set can be found in Appendix A1.

4.4.2 EDRM

The EDRM file format data set was downloaded from the EDRM website (57). This dataset contains 381 files covering 200 file formats and is designed for use in e-discovery tests. The dataset comes with a reference set listing information about each file. During our experiments, we discovered that this reference set provided by EDRM was not up to date with regards to the actual content of the dataset. To remedy this, we made our own reference set to be used for result comparison.

The EDRM data set was downloaded from the EDRM website as an archive. The archive was unpacked and the files transferred to each of the USB drives formatted as *NTFS*, *FAT32*, *EXT4*, and *HFS+*. The MD5 checksum of the downloaded zip archive is 105e46afd5a65c524a2fa7ff05b5d580.

The reference set for the data set is listed in Appendix A2.

4.4.3 Images and videos

The purpose of this dataset is to test if a digital forensic tool is able to list and extract active and deleted images and videos of various formats. The images and videos used in the dataset are gathered from the NIST CFReDS project (62). These datasets contain various graphic and video formats.

In addition to the files downloaded from NIST, we added a photo and a video from a mobile phone with geotagging activated[8]. This was done to test Deepthought's parsing of EXIF data stored in the image and video. EXIF information is additional information

[8] Geotagging adds information about the GPS location where the image was taken.

stored in the photo, often listing the camera make and model, timestamps of when the image was taken, and GPS coordinates if available. Two images and videos of each file type were renamed with the format *deleted-filename* to indicate which files to manually delete on all file systems after being transferred.

The reference set for the dataset is listed in Appendix A3 – Graphics and video.

4.4.4 Email

According to the user manual (37), Deepthought supports *dbx*, *eml*, *msg*, *ost*, *mbox*, and *pst* email formats. To generate the email dataset, we created three email accounts with three differend email providers. One on Gmail; one on Yahoo; and one on Outlook. We created content by sending emails between these accounts, and downloaded the email archives.

The Gmail email archive was downloaded using the built-in google takeout feature in Gmail. This extracts the email archive as a *mbox* file. Three of the emails in the Gmail account were downloaded separately as *eml* files using the built-in *save as* feature.

The Yahoo and Outlook accounts were downloaded using Microsoft Outlook 2016 and connecting to the accounts. The Yahoo email account was extracted as the original *ost* file type, and the Outlook email account was extracted by exporting the archive as a *pst* file.

The *msg* files were copied from the EDRM dataset used in test 2. We were unable to create a *dbx* archive because this file type is used by Microsoft Outlook Express, which is discontinued.

The reference set for the dataset is listed in Appendix A4.

4.4.5 Archives

Deepthought should support extracting images from archives, and extracting encrypted archives larger than 500MB. Deepthought should also support extracting documents but we don't know if Deepthought is able to extract documents from archives. The unencrypted archives were made using 7zip version 19.00.

The dataset consists of 6 unencrypted archives of various formats, each containing one image from the *images and videos* dataset used in test 3(41). The unencrypted archive formats containing images are *zip*, *tar*, *tar.gz*, *tar.bz2*, *wim* and *7z*. In addition, the dataset has 5 unencrypted archives of various formats, each containing a PDF file, a Word document, and an Excel document. The unencrypted archives containing documents are *zip*, *tar*, *tar.gz*, *wim* and *7z*.

There are 8 encrypted archives of the following formats: *zip*, *7z*, *bitlocker*, and *Veracrypt*. There are two of each format, one file smaller than 50MB and one larger than 500MB. This is to test if Deepthought is able to extract all the large archives, and will not export the smaller archives.

The reference set for the dataset can be found in Appendix A5.

4.4.6 Documents

To test the image extraction feature from documents, we created a set of common document types: *ODT*, *ODS*, *docx*, *xlsx*, *pptx*, *rtf*, and *PDF*. 5 images from the NIST CFReDS dataset (62) described in Section 4.4.3 were edited in Microsoft Paint, adding a text field showing what document type each image originated from, and were added to

the corresponding documents. This was done to create a unique MD5 for each image for easier comparison when analyzing the results later. This also made it easier to visually confirm where the image was extracted from, in case the MD5 of the extracted images did not match.

Each of the documents contained a *jpg*, *bmp*, *gif*, *png* and a *tiff* image.

The reference set can be found in Appendix A6.

4.4.7 Browser history

When researching datasets, we were unable to find any datasets with browser history from multiple browsers across multiple operating systems, so we decided to generate this dataset ourselves. We selected Google Chrome, Mozilla Firefox, Safari, and Microsoft Edge as the browsers to generate the history files. These browsers were selected because they are some of the most common browsers on the three operating systems.

The processes of generating web history logs consisted of visiting 4 websites, doing 4 Google searches, and downloading one image file. The URLs and search words are listed in Table 3. The URL used for downloading the image forces the browser to download the image. This is not necessarily logged as a visited webpage by all browsers, and will be commented if encountered in our tests. The process was repeated for each operating system in case the browsers generate history files in different formats depending on the operating system. Table 4 lists the browsers used on the three operating systems.

To generate the reference set, we recorded the timestamps for each visit to the websites. The timestamps were recorded using browser extensions in Firefox and Chrome, and manually in Microsoft Edge and Safari. The browser extension used was *History Master* version 2.1.4 by Jiacai Liu for Chrome and Firefox. The databases containing the browsing history for each of the browsers were extracted and included in the dataset.

The reference set of recorded timestamps of each browser can be found in Appendix A7.

Activity	**URL / Word**
Visit URL	https://www.nist.com
Visit URL	https://www.swgde.org/
Visit URL	https://www.edrm.net/
Visit URL	Google.com
Search for word	Digital
Search for word	Forensics
Search for word	Computer
Search for word	Software
Download	goo.gl/FrUcJM[9]

Table 3 - Browser activity.

[9] goo.gl/FrUcJM is a shortened URL which redirects to https://unsplash.com/photos/p0j-mE6mGo4/download?force=true. Opening this link will download a free image by Lorenzo Herrera from unsplash.com.

The browsers selected per operating system are listed in Table 4.

	Microsoft Edge	**Google Chrome**	**Mozilla Firefox**	**Safari**
Windows	X	X	X	
Ubuntu Linux		X	X	
OSX		X	X	X

Table 4 - Browser distribution on operating systems.

4.5 Test environment

The datasets generated for this thesis were created on three laptops running different operating systems. After the tests were executed, the results were analyzed on a separate machine with better specifications for running the processing and analysis tools.

4.5.1 OSX environment

Description	**Type**
Laptop	HP Elitebook
Harddrive	Hitatchi HTS727550A9E364 500GB
Operating system	OSX Mountain Lion
Browsers	• Safari v. 11.0.1 • Firefox v. 65.0.2 • Google Chrome v. 72.0.0.3626
Browser extensions	• Chrome: History master version 2.1.4 by jiacai2050. • Firefox: History master version 2.1.4 by Jiacai Liu

Table 5 - System setup, OSX environment.

4.5.2 Linux environment

Description	Type
Laptop	HP Elitebook
Harddrive	Seagate ST9500420ASG 500GB
Operating system	Ubuntu 18.04.2 LTS
Browsers	• Firefox v. 65.0.2 • Google Chrome v. 72.0.0.3626.109
Browser extensions	• Chrome: History master version 2.1.4 by jiacai2050. • Firefox: History master version 2.1.4 by Jiacai Liu

Table 6 - System setup, Linux environment.

4.5.3 Windows environment

Description	Type
Laptop	HP Elitebook
Harddrive	HGST H2T5003272S7 500GB
Operating system	Windows 10
Browsers	• Firefox v. 65.0.2 • Google Chrome v. 72.0.0.3626.109 • Microsoft Edge v. 42.17134.1.0
Browser extensions	• Chrome: History master version 2.1.4 by jiacai2050. • Firefox: History master version 2.1.4 by Jiacai Liu

Table 7 - System setup, Windows environment.

4.5.4 Analysis machine

Description	Type
Type	Macbook Pro A1707
CPU	Intel Core i7 2,9 GHz
RAM	16 GB 2133 MHz LPDDR3
Storage	APPLE SSD SM1024L 1TB harddrive split into two 500GB partitions. One partition for OSX, one for Windows 10.
Operating system	Windows 10 PRO build 17134 (Bootcamp)
Forensic software	• FTK Imager 4.2.0.13 • FTK Lab 7.0 • Microsoft Office 2016 • Veracrypt 1.23-Hotfix-2 • 7zip version 19.00 (64)

Table 8 - System setup, analysis machine.

4.5.5 Write-blocker

A write-blocker is used for preserving the integrity of connected storage devices by blocking all commands making changes to the device. The write-blocker used in our experiments is a Logicube WriteProtect Desktop.

4.6 Reference tool and reference sets

It is not always possible to generate a reference set covering all aspects of the dataset without using other digital forensic tools. This becomes apparent when dealing with embedded files like images in a document. One could include the original image in the reference set, but if the process of embedding the image changes the image in any way, the reference set would no longer match the dataset used as input. To remedy this, we process the forensic images in the digital forensic tool AD Lab[10], in addition to comparing against the original reference set.

As described by Friheim, commercial tools can also give incorrect output (64). Best practice would be to use more than one tool to generate the data for comparing results, but because of time restrictions we will only use one tool for comparing results.

We will use AD Lab 7.0 to verify the results produced by Deepthought. AD Lab 7.0 is a well-known investigative platform for full-scale digital investigations (34).

The reference sets generated from the original files were created using the free software *md5Deep64* version 4.3[11]. The individual reference sets can be found in Appendix A1 – A7.

The reference sets, and reference tool output were also copied into the Microsoft Excel spreadsheets together with the results from Deepthought for semi-automatic comparison of data.

4.7 Combining results

Deepthought presents its findings as a web page with multiple tabs listing the artifacts. With datasets containing more than just a few files, we need to introduce some form of automation for comparing the results produced by Deepthought, the reference tool and reference set. This could be done by writing a script to extract the data from the web report and comparing them to the results produced by AD Lab and reference set. We chose not to use this technique because any future changes to the web report could result in the script no longer working, reducing the reusability of the script.

We opted for manually copying the data from the relevant tabs in the web report, into Microsoft Excel sheets. Microsoft Excel is a great tool with scripting[12] functionality made for listing and comparing data. This functionality was used to compare MD5 checksums

[10] AD Lab was recently renamed from FTK Lab.

[11] MD5Deep64 is a open source tool for calculating MD5 checksums of files, and is available from http://md5deep.sourceforge.net/

[12] By scripting in this context we refer to the formula functions in Excel. Excel does also support much more advanced scripting language called Visual Basic for Applications (VBA).

and timestamps. The results from all the tests are listed in Appendix B1 through Appendix B30, and will be shared as a zip archive rather than printed as part of the thesis. This is because the formats of the Excel sheets are too inconvenient to be added as printed pages.

4.8 Deepthought settings

Deepthought lets the investigator select parsing options to use when processing a device. Throughout all our experiments we will run processing with all options selected. This is done to remove potential errors from not running all modules of Deepthought.

Deepthought has two levels of processing files, *live and deleted files only*, and *live/deleted/unallocated*. Because we do not know how Deepthought processes deleted files, we will use the *live/deleted/unallocated* setting when processing evidence with deleted files. For all experiments where all files are active, we will use the *live and deleted files only* setting.

5 Experiments and results

In this chapter, we will detail each of the tests conducted to verify the features outlined in Chapter 3.

5.1 Test 1 – DFR-01 Recover one deleted non-fragmented file

5.1.1 Scope

The scope of this test is to determine if Deepthought is able to list active files and a deleted non-fragmented file from the *NTFS, FAT, Ext, ExFAT* and *HFS+* filesystems. The DFR-01 test is designed by NIST CFTT (50). More information about the dataset is detailed in Chapter 4.

5.1.2 Assertions

- Deepthought should list all active files.
- Deepthought should display correct metadata for all discovered active files.
- Deepthought should discover all deleted files.
- Deepthought should display correct metadata for all discovered deleted files.

5.1.3 Test execution

The DFR-01 forensic images for each file system were downloaded from the NIST CFReDS website (41) and transferred to the Freetool distribution using a Kingston DataTraveler G4 8GB USB flash drive. Each of the forensic images were processed using Deepthought, launched from the desktop, with all features activated. The option of processing *Live, Deleted and unallocated* was selected, and the output of Deepthought was stored locally on the same storage as the operating system. We used the *Live, Deleted and unallocated* setting because this test involves deleted files, and we assume this setting is optimized for detecting and recovering deleted files. The DFR-01 forensic images were then processed in the reference tool AD Lab 7.0. For this test, there was no reference set containing MD5s provided by NIST.

5.1.4 Results

Criteria	**NTFS**	**FAT**	**EXT**	**HFS+**	**ExFAT**
List all active files.	Pass	Pass	Pass	Fail	Pass
Display correct metadata for all discovered active files.	Fail	Fail	Fail	Fail	Fail
List all deleted files.	Fail	Fail	Fail	Fail	Fail
Discover and display correct metadata for all deleted files	Fail	Fail	Fail	Fail	Fail

Table 9 - Test 1 results.

NTFS:
Deepthought did not list the deleted file.

In the NTFS tests, Deepthought listed created timestamps that were 2 minutes off the timestamps produced by AD Lab. Deepthought lists the same modified timestamps as AD Lab.

FAT:
Deepthought did not list the deleted files.

The modified timestamps listed by Deepthought are the same as listed by AD Lab. Deepthought does not list any created timestamps, but lists accessed timestamps. AD Lab does not list any accessed timestamps, but is listing the created timestamps. The accessed timestamps listed by Deepthought do not match the created timestamps listed by AD Lab, indicating that the timestamps are not simply a mislabeled column.

EXT4:
Deepthought did not list any deleted files. AD Lab was only able to fully list and recover one of the three deleted files, and was unable to recover the file name of the deleted file.

Deepthought lists the same modified timestamps as AD Lab. Deepthought lists created timestamps for all active files. AD Lab does not list created timestamps for the files. The accessed timestamps listed by Deepthought match the accessed timestamps listed by AD Lab.

HFS+:
Deepthought lists no information about files stored on the HFS filesystems.

ExFAT:
Deepthought did not list any deleted files. AD Lab was only able to recover metadata of the deleted files.

The modified and accessed timestamps listed by Deepthought are 3 hours off the timestamps listed by AD Lab. Deepthought does not list created timestamps.

General findings:
The files in the DFR-01 dataset from NIST contains only text files. The results in this test could indicate that Deepthought does not search for all deleted files, but might be looking for specific file types.

Complete results are listed in Appendix B1 – B5.

5.2 Test 2 – EDRM Various file format extraction

5.2.1 Scope

Determine if Deepthought is able to recover files of various formats from *NTFS, FAT32, Ext4,* and *HFS+* filesystems. More information about the dataset is detailed in Chapter 4. By running this public dataset we will test multiple features of Deepthought; Active file listing, image extraction, document extraction, archive listing, and email parsing.

5.2.2 Assertions

- Deepthought should list all files in the reference set.
- Deepthought should list the correct timestamps for all the files in the reference set.
- Deepthought should extract all jpg files in the reference set.
- Deepthought should extract all gif files in the reference set.
- Deepthought should extract all bmp files in the reference set.
- Deepthought should display all identified images in the *Image* tab.
- Deepthought should extract all doc files.
- Deepthought should extract all xls files.
- Deepthought should extract all ppt files.
- Deepthought should extract all msg emails.
- Deepthought should extract all pst emails.
- Deepthought should extract all jpg images from archives.
- Deepthought should extract all images from PDFs.

5.2.3 Test execution

The EDRM dataset was transferred to four Kingston DataTraveler G4 32GB USB flash drives formatted as NTFS, FAT32, EXT4, and HFS+ filesystems. Each of the flash drives were processed using Deepthought, launched from the shortcut on the desktop of the Freetool distribution. The processing was executed, using all features of Deepthought, and only processing *live and deleted* files (E.g. not unallocated). Unallocated was not selected for this test because there are no deleted files in the dataset.

All images and videos identified by the software were manually selected and tagged as *Child Explicit* in the file browser provided by Deepthought. This was done for easier review in the web report later.

After processing we navigated to the folder we had selected for the processed output, and made lists of MD5 checksums of each of the extracted files[13]. After calculating the checksums, we opened the web report and documented all tabs containing information using the Firefox extension *Fireshot* version 0.98.96 by Susbox, which comes pre-installed with the Freetool distribution.

By searching for an empty string in the tab *File Search by MD5 Value* it returns all files found by Deepthought. The *File Search by MD5 Value* and *Images* tab were saves as html because Fireshot has limited support for long web pages. Saving these tabs to html also made it easier to transfer the information to excel for analysis.

[13] MD5 checksums were made using the included software md5deep version 4.4 using the following command scheme as root user:

md5deep –r {path_to_folder_source_folder} > {filename_for_list_of_md5s.txt}

After documenting the results, we disconnected the USB and made a forensic copy for before and after comparison in test 7. The output from Deepthought was transferred to a different computer for analysis, and the forensic image made before running Deepthought was processed using FTK Imager.

5.2.4 Results

Criteria	**NTFS**	**FAT32**	**EXT4**	**HFS+**
List all files in the reference set.	Pass	Pass	Pass	Fail
List the correct timestamps for all the files in the reference set.	Fail	Fail	Pass	Fail
Extract all jpg files in the reference set.	Pass	Pass	Pass	Fail
Extract all gif files.	Pass	Pass	Pass	Fail
Extract all bmp files.	Fail	Fail	Fail	Fail
Display all identified images in the *Image* tab.	Fail	Fail	Fail	Fail
Extract all doc files.	Pass	Pass	Fail	Fail
Extract all xls files.	Pass	Pass	Pass	Fail
Extract all ppt files.	Pass	Pass	Pass	Fail
Extract all msg emails.	Fail	Fail	Fail	Fail
Extract all pst emails.	Pass	Pass	Pass	Fail
Extract jpg image from zip archive.	Pass	Pass	Pass	Fail

Table 10 - Test 2 results.

General findings:
On all filesystems, Deepthought extracted all the images except one. This error carries over to the criteria that all identified images should be listed in the image tab. Deepthought recovered more files than expected.

71 images were recovered in total, in each of the tests. This shows that Deepthought supports recovery of the following image filetypes: *bmp*, *dcx*, *emf*, *g01*, *gif*, *hwp*, *ico*, *img*, *jpg*, *pbm*, *pcx*, *pgm*, *png*, *ppm*, *ras*, *tga*, *tif*, *wk3*, *wmf*, *xwd*. All files listed in the image tab were extracted.

Deepthought was unable to extract or parse *msg* email files. Folders for each *msg* file were created but all the folders were empty.

Deepthought did not extract the *pst* email archive in its original form, but parsed its content into separate files. The folder structure was preserved, and the parsed content was correct. None of the parsed email content was displayed in the web report.

All the expected images except one were extracted correctly on all filesystems. The image Deepthought was unable to extract, was extracted correctly by AD Lab, indicating there was nothing wrong with the image. All documents were extracted correctly by Deepthought.

Deepthought reported and extracted a few false positives. Deepthought successfully extracted a jpg image from a zip archive, but it also extracted the zip archive containing the image as a jpg image file. Deepthought extracted a folder containing a PDF as a PDF file.

The *Modified* timestamps listed are correct for all files. All *created* and *accessed* timestamps listed in the FAT32 file system are incorrect, compared to the AD Lab results. In the EXT4 file system Deepthought listed all timestamps correctly. In the NTFS file system, Deepthought listed different created timestamps than AD Lab in 382 out of 392 files.

Complete results are listed in Appendix B6 – B9.

5.3 Test 3 – Image and video

5.3.1 Scope

Determine if Deepthought can detect and extract images and videos from *NTFS, FAT32, EXT4* and *HFS+* file systems.

5.3.2 Assertions

- Deepthought should detect and extract all non-deleted jpg images.
- Deepthought should detect and recover all deleted jpg images.
- Deepthought should present all identified jpg images in the image tab.
- Deepthought should detect and extract all non-deleted png images.
- Deepthought should detect and recover all deleted png images.
- Deepthought should present all identified png images in the image tab.
- Deepthought should detect and extract all non-deleted bmp images.
- Deepthought should detect and recover all deleted bmp images.
- Deepthought should present all identified bmp images in the image tab.
- Deepthought should detect and extract all non-deleted gif images.
- Deepthought should detect and recover all deleted gif images.
- Deepthought should present all identified gif images in the image tab.
- Deepthought should detect and extract all non-deleted tiff images.
- Deepthought should detect and recover all deleted tiff images.
- Deepthought should present all identified tiff images in the image tab.
- Deepthought should detect and extract all non-deleted mp4 videos.
- Deepthought should detect and extract all deleted mp4 videos.
- Deepthought should present all identified mp4 videos in the movies tab.
- Deepthought should detect and extract all non-deleted MOV videos.
- Deepthought should detect and extract all deleted MOV videos.
- Deepthought should present all identified MOV videos in the movies tab.
- Deepthought should detect and extract all non-deleted avi videos.
- Deepthought should detect and extract all deleted avi videos.
- Deepthought should present all identified avi videos in the movies tab.
- Deepthought should detect and extract all non-deleted ogv videos.
- Deepthought should detect and extract all deleted ogv videos.
- Deepthought should present all identified ogv videos in the movies tab.
- Deepthought should detect and extract all non-deleted 3gp videos.
- Deepthought should detect and extract all deleted 3gp videos.
- Deepthought should present all identified 3gp videos in the movies tab.
- Deepthought should detect and extract all non-deleted wmv videos.
- Deepthought should detect and extract all deleted wmv videos.
- Deepthought should present all identified wmv videos in the movies tab.
- Deepthought should list correct metadata for all identified non-deleted files.
- Deepthought should list correct metadata for all identified deleted files.
- Deepthought should list correct EXIF data for the image and video containing geo location.

5.3.3 Test execution

The files containing the browser history were transferred to four Kingston DataTraveler G4 8GB USB memory sticks, formatted as each of the file systems. The memory sticks were then connected to a Logicube WriteProtect Desktop write-blocker and acquired as a forensic image using FTK Imager 4.2.0.13.

Each of the flash drives were then connected to the Freetool distribution and processed using Deepthought. The processing was executed, using all features of Deepthought, and the option of processing *Live and deleted and Unallocated*.

We used the *Live, Deleted and unallocated* setting because this test involves deleted files, and we assume this setting is optimized for detecting and recovering deleted files. The output of Deepthought was stored locally on the storage hosting the operating system.

All images and videos identified by the software were manually selected and tagged as *Child Explicit* in the file browser provided by Deepthought. This was done for easier review in the web report later.

After processing, MD5 checksums were calculated for each of the extracted files using *Md5deep*. Next, the web report was opened and all tabs with information were documented using the Firefox extension *Fireshot* version 0.98.96 by Susbox.

After being processed by Deepthought, the memory sticks were connected to the Logicube write-blocker once more and made a second forensic copy of each device, using FTK Imager 4.2.0.13. The forensic images were then processed in the reference tools AD Lab 7.0 for comparing results.

5.3.4 Results

Criteria	**NTFS**	**FAT32**	**EXT4**	**HFS+**
Detect and extract all non-deleted jpg images.	Pass	Pass	Pass	Fail
Detect and recover all deleted jpg images.	Pass	Pass	Pass	Fail
Present all identified jpg images in the image tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted png images.	Pass	Pass	Pass	Fail
Detect and recover all deleted png images.	Pass	Pass	Pass	Fail
Present all identified png images in the image tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted bmp images.	Pass	Pass	Pass	Fail
Detect and recover all deleted bmp images.	Fail	Fail	Pass	Fail
Present all identified bmp images in the image tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted gif images.	Pass	Pass	Pass	Fail
Detect and recover all deleted gif images.	Fail	Fail	Pass	Fail
Present all identified gif images in the image tab.	Fail	Pass	Fail	Fail
Detect and extract all non-deleted tiff images.	Pass	Pass	Pass	Fail
Detect and recover all deleted tiff images.	Fail	Fail	Pass	Fail
Present all identified tiff images in the image tab.	Pass	Pass	Fail	Fail
Detect and extract all non-deleted mp4 videos.	Pass	Pass	Pass	Fail

Detect and extract all deleted mp4 videos.	Fail	Fail	Pass	Fail
Present all identified mp4 videos in the movies tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted MOV videos.	Pass	Pas	Pass	Fail
Detect and extract all deleted MOV videos.	Fail	Fail	Pass	Fail
Present all identified MOV videos in the movies tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted avi videos.	Pass	Pass	Pass	Fail
Detect and extract all deleted avi videos.	Fail	Fail	Pass	Fail
Present all identified avi videos in the movies tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted ogv videos.	Pass	Pass	Pass	Fail
Detect and extract all deleted ogv videos.	Fail	Fail	Pass	Fail
Present all identified ogv videos in the movies tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted 3gp videos.	Pass	Pass	Pass	Fail
Detect and extract all deleted 3gp videos.	Fail	Fail	Pass	Fail
Present all identified 3gp videos in the movies tab.	Fail	Fail	Fail	Fail
Detect and extract all non-deleted wmv videos.	Pass	Pass	Pass	Fail
Detect and extract all deleted wmv videos.	Fail	Fail	Pass	Fail
Present all identified wmv videos in the movies tab.	Fail	Fail	Fail	Fail
List correct metadata for all identified non-deleted files.	Fail	Fail	Pass	Fail
List correct metadata for all identified deleted files.	Fail	Fail	Fail	Fail
List correct EXIF data for the image and video containing geo location.	Pass	Pass	Pass	Fail

Table 11 - Test 3 results.

General findings:

Deepthought consistently extracted the same deleted video files with the same erroneous MD5 checksums on all filesystems. This indicates that there might be something wrong with the method of extracting the deleted videos.

For all tests, Deepthought did not list any movies in the movies tab of the web report. This feature does not seem to be working as intended. The EXIF tab in the web report only lists geo location EXIF information for the image, not for the video.

Across the tests of all filesystems Deepthought is listing different created timestamps than AD Lab.

NTFS:

Deepthought was able to detect and export all deleted *png* and *jpg* files, but was unable to list or export any of the deleted *gif*, *tiff* or *bmp* files.

Deepthought was unable to recover any filenames from the deleted video and image files. Deepthought recovered 32 of 37 video files, 6 of which had different MD5 checksums than the reference set, and AD Lab. We were unable to play three of the recovered video files with different MD5 checksums.

Deepthought listed and extracted two bmp images as encrypted files. As a result, these files were not listed in the images tab.

FAT32:
All created timestamps listed by Deepthought are incorrect. Deepthought was unable to recover filenames of the recovered deleted files.

In the FAT32 filesystem, Deepthought reports *Created* timestamp as empty, and has populated *Accessed* timestamps. AD Lab reports that the *Created* timestamps are populated and the *Accessed* timestamp are empty.

Deepthought failed to export 11 deleted videos.

EXT4:
Deepthought extracted all active and deleted images in the EXT4 file system, but was unable to recover the filenames of the deleted files. Two of the bmp images were extracted as encrypted archives, but the MD5 of the extracted files match the reference set.

All active and deleted movies were extracted, but Deepthought was unable to recover filenames of the deleted files. AD lab recovered all filenames for the deleted files. The image tab did not list any of the identified images for the EXT4 file system.

All modified timestamps listed are the same as the timestamps listed by AD Lab, but 57/100 created timestamps did not match the created timestamps from AD Lab.

HFS+:
Deepthought lists no information about files stored on the HFS+ filesystem.

Complete results are listed in Appendix B10 – B13.

5.4 Test 4 – Email

5.4.1 Scope

Determine if Deepthought can extract email archives from *NTFS, FAT32, Ext4,* and *HFS+* filesystems. More information about the dataset is detailed in Chapter 4.

5.4.2 Assertions

- Deepthought should extract all mbox emails.
- Deepthought should extract all pst emails.
- Deepthought should extract all ost emails.
- Deepthought should extract all msg emails.
- Deepthought should extract all eml emails.

5.4.3 Test execution

The email archives were transferred to 4 Kingston DataTraveler G4 8GB USB memory sticks formatted as each of the defined file systems. The memory sticks were then connected to a Logicube WriteProtect Desktop write-blocker and acquired as a forensic image using FTK Imager 4.2.0.13.

Each of the memory sticks were then connected to the stand alone Freetool distribution and processed using Deepthought. The evidence was processed using all features activated, and only processing *live and deleted* files (E.g. not unallocated). We used the *Live, Deleted files only* setting because this test does not involve deleted files.

The output of Deepthought was stored locally on the storage hosting the operating system and transferred to a separate machine for further analysis.

After processing, MD5 checksums were calculated for each of the extracted files using *Md5deep*. Next, the web report was opened and all tabs with information were documented using the Firefox extension *Fireshot* version 0.98.96 by Susbox.

After being processed by Deepthought, the memory sticks were connected to the Logicube write blocker and a second forensic copy was made of each device, using FTK Imager 4.2.0.13. The forensic images were then processed in the reference tools AD Lab 7.0 for comparing results. Because the extracted emails produced by Deepthought are files generated by the parsers used to analyze the email archives, the parsed files should be the same across all file systems. The outputted files from *NTFS* filesystem were manually verified against AD Lab once, and then the MD5s of the verified files were compared to the parsed files on each of the other filesystems. This way we could quickly repeat the verification process, and conduct further analysis if any of the files did not match the verified files.

5.4.4 Results

Criteria	NTFS	FAT32	Ext4	HFS+
Extract all mbox emails.	Pass	Pass	Pass	Fail
Extract all pst emails.	Pass	Pass	Pass	Fail
Extract all ost emails.	Pass	Pass	Pass	Fail
Extract all msg emails.	Fail	Fail	Fail	Fail
Extract all eml emails.	Pass	Pass	Pass	Fail

Table 12 - Test 4 results.

General findings:
Deepthought is able to extract the full *mbox* archive. Folders are created next to the extracted file, but the folders are not populated by parsed emails. We regard these results to pass the test because the archive is extracted correctly in its original format.

Deepthought is able to parse the *pst* email archive, and extracts each email by storing the headers and message body into separate files. The information stored in each file is correct and the folder structure is preserved. Deepthought passes *pst* archive test because all the information from the parsed emails is present and correct.

Deepthought does not extract the full *ost* archive in its original format, but parses all emails in the archive correctly. Emails are stored with the headers and message body in separate files, and the folder structure is intact. Deepthought passes the *ost* archive test because all the information from the emails is present in the parsed files.

Deepthought creates a folder for each *msg* email, but they are empty. Deepthought is unable to extract or parse *msg* emails.

Complete results are listed in Appendix B14 – B17.

5.5 Test 5 – Archives and encrypted files

5.5.1 Scope

The scope of this test is to determine if Deepthought can detect and display archives and encrypted files from *NTFS, FAT32, Ext4,* and *HFS+* filesystems. The test is also designed to verify if Deepthought is able to extract images and documents from the archives. More information about the dataset is detailed in Chapter 4.

5.5.2 Assertions

- Deepthought should list all archives in the reference set.
- Deepthought should detect and list correct metadata for all the listed files.
- Deepthought should extract all images from the archives containing images.
- Extracted images should be listed in the images tab.
- Deepthought should detect and extract all encrypted archives over 500MB
- Deepthought should not extract archives smaller than 500MB.
- All encrypted archives should be listed in the encrypted tab.

5.5.3 Test execution

The archives and encrypted files were transferred to Kingston DataTraveler G4 8GB USB memory sticks formatted as each of the file systems. The memory sticks were then connected to a Logicube WriteProtect Desktop write-blocker and acquired as a forensic image using FTK Imager 4.2.0.13.

Each of the memory sticks were then connected to the Freetool distribution and processed using Deepthought. All features were activated for processing, and Deepthought was set to only parse *active and deleted* files.

After processing, MD5 checksums were calculated for each of the extracted files using *Md5deep*. Next, the web report was opened and all tabs with information were documented using the Firefox extension *Fireshot* version 0.98.96 by Susbox.

The output of Deepthought was stored locally on the storage hosting the operating system.

After being processed by Deepthought, the memory sticks were connected to the Logicube write-blocker once more and made a second forensic copy of each device, using FTK Imager 4.2.0.13. The forensic images were then processed in the reference tools AD Lab 7.0 for comparing results.

5.5.4 Results

Criteria	NTFS	FAT32	EXT4	HFS+
List all archives in the reference set.	Pass	Pass	Pass	Fail
Detect list correct metadata for all the listed files.	Fail	Fail	Fail	Fail
Extract all images from the archives containing images.	Pass	Pass	Fail	Fail
Extracted images should be listed in the images tab.	Fail	Fail	Fail	Fail
Detect and extract all encrypted archives over 500MB	Fail	Fail	Fail	Fail
Not extract archives smaller than 500MB.	Fail	Fail	Fail	Fail

Table 13 - Test 5 results.

General findings:
The results in this test were the same across all file systems except HFS+. Deepthought produced no information about any of the files on the HFS+ filesystem.

Deepthought was able to list all files in the reference set for the other filesystems.

Deepthought detected all the images contained in the unencrypted archives, but only extracted one image. Deepthought stores images as jpg files, using the MD5 of the image as filename. Because all archives contained the same image, this way of storing the extractions makes it impossible to determine if the image was extracted once, or if it was extracted for each of the archives, overwriting the previous extraction. The image was not listed in the Images tab.

Deepthought erroneously exported one of the archives as an image file.

Only two of the encrypted archives were extracted. One of the extracted archives was less than 500MB.

Because Deepthought supports document extraction, the dataset contained archives containing documents. None of the documents were detected or extracted by Deepthought.

NTFS:
Deepthought listed all images in the archives. All modified timestamps match the modified timestamps reported by AD Lab, except for the image files detected inside the archives. None of the created or accessed timestamps matched the AD Lab results.

FAT32:
All the files had different timestamps than reported by AD Lab. The modified timestamps were in most cases just a few seconds off, but the created and accessed timestamps had a larger time difference. If the timestamps were just one second off, we would assume it was due to a rounding error. However, because many of the timestamps are off by 2 seconds we believe it is an error in how Deepthought interprets the timestamps.

EXT4:
For all files, the accessed timestamps reported by Deepthought matches the created time listed by AD Lab. The Created timestamp listed by Deepthought seems to be a copy of the accessed time reported by Deepthought. All modified timestamps matched the timestamps reported by AD Lab.

Deepthought was unable to list the images inside the *wim*, *zip* and *tar.gz* archives. Because of this, Deepthought failed the image extraction test.

HFS+:
Deepthought lists no information about files stored on the HFS+ filesystem.

Complete results are listed in Appendix B18 – B21.

5.6 Test 6 – Documents

5.6.1 Scope

Determine if Deepthought is able to extract all the document formats it should support, from *NTFS, FAT32, Ext4,* and *HFS+* filesystems. The test is also designed to test if Deepthought is able to extract all images from PDF documents, and see if it supports image extraction from other document formats.

5.6.2 Assertions

- Deepthought should extract all docx files.
- Deepthought should extract all pptx files.
- Deepthought should extract all xlsx files
- Deepthought should extract all ods files.
- Deepthought should extract all odt files.
- Deepthought should extract all PDF files
- Deepthought should extract all images from the PDF document.

5.6.3 Test execution

The documents were transferred to Kingston DataTraveler G4 8GB USB flash drives formatted as each of the file systems. The USB flash drives were then connected to a Logicube WriteProtect Desktop write-blocker and acquired as a forensic image using FTK Imager 4.2.0.13.

Each of the memory sticks were then connected to the Freetool distribution and processed using Deepthought. All features described in Section 2.5 were activated for processing, and Deepthought was set to only parse *active and deleted* files. Unallocated was not selected because there are no deleted files in the dataset.

The output of Deepthought was stored locally on the storage hosting the operating system.

After processing, MD5 checksums were calculated for each of the extracted files using *md5deep*. Next, the web report was opened and all tabs with information were documented using the Firefox extension *Fireshot* version 0.98.96 by Susbox.

After being processed by Deepthought, the USB flash drives were connected to the Logicube write blocker and a second forensic copy of each device was made using FTK Imager 4.2.0.13. The forensic images were then processed in the reference tools AD Lab 7.0 for comparing results.

5.6.4 Results

Criteria	NTFS	FAT32	EXT4	HFS+
Extract all docx files.	Pass	Pass	Pass	Fail
Extract all pptx files.	Pass	Pass	Pass	Fail
Extract all xlsx files	Pass	Pass	Pass	Fail
Extract all ods files.	Pass	Pass	Pass	Fail
Extract all odt files.	Pass	Pass	Pass	Fail
Extract all PDF files	Pass	Pass	Pass	Fail
Extract all images from the PDF document.	Fail	Fail	Fail	Fail

Table 14 - Test 6 results.

Greneral findings:

On all filesystems except HFS+, Deepthought extracted all the documents correctly. Deepthought should support extracting images from PDF documents, but was only able to recover one of the five images embedded in the PDF document. The extracted image did not have the same MD5 as the MD5 in the reference set. All documents had five images, but Deepthought did not recover any images from any of the other document formats. This corresponds with the user manual (37) which only specifies image extraction from PDF.

Complete results are listed in Appendix B22 – B25.

5.7 Test 7 – Browser History

5.7.1 Scope

Determine if Deepthought can detect, parse and display browser history from Safari, Firefox, Microsoft Edge and Google Chrome, from *NTFS, FAT32, Ext4,* and *HFS+* filesystems. More information about the dataset is detailed in Chapter 4.

5.7.2 Assertions

- Deepthought should display the correct browser history from Google Chrome.
- Deepthought should display the correct the browser history from Mozilla Firefox.
- Deepthought should display the correct the browser history from Microsoft Edge.
- Deepthought should display the correct the browser history from Safari.

5.7.3 Test execution

The files containing the browser history were transferred to 4 Kingston DataTraveler G4 8GB USB flash drives, formatted as each of the file systems. The memory sticks were then connected to a Logicube WriteProtect Desktop writeblocker and acquired as a forensic image using FTK Imager 4.2.0.13.

Each of the USB flash drives were then connected to the stand alone Freetool distribution and processed using Deepthought with all features activated. The output of Deepthought was stored locally on the storage hosting the operating system.

After being processed by Deepthought, the memory sticks were connected to the Logicube write-blocker once more and made a second forensic copy of each device, using FTK Imager 4.2.0.13. The forensic images were then processed in the reference tool AD Lab 7.0 for comparing results.

5.7.4 Results

Criteria	**NTFS**	**FAT32**	**EXT4**	**HFS+**
Display correct Google Chrome browser history.	Fail	Fail	Fail	Fail
Display correct Mozilla Firefox browser history.	Fail	Fail	Fail	Fail
Display correct Microsoft Edge browser history.	Fail	Fail	Fail	Fail
Display correct Safari browser history	Fail	Fail	Fail	Fail

Table 15 - Test 7 results.

General findings:

Deepthought listed all files in the reference set on all file systems. None of the browser histories were displayed in the web report. None of the databases containing the browser history were exported. This indicates that this feature does not work as intended.

Complete results are listed in Appendix B26 – B29.

5.8 Test 8 – Forensic soundness

5.8.1 Scope

Test if running Deepthought affects the storage device being investigated.

5.8.2 Assertions

Running Deepthought should not affect the storage device being investigated.

All MD5 checksums of the forensic images should match the MD5 checksums of the images calculated after running Deepthought.

5.8.3 Test execution

In test 2 through test 7, the USB flash drives were connected to a write-blocker device[14] and forensically imaged using the commercial software FTK Imager. This process creates a forensic copy of the device, and calculates the MD5 checksum of the image. This process was repeated after each test. By comparing the MD5 checksums calculated before and after each test, we will be able to see if running Deepthought affected the device in any way.

The DFR-01 datasets used in test 1 were not included in the test, because these datasets come as premade forensic images.

After executing test 6 we made an error, and formatted the USB drives before creating the forensic copies. It was therefore impossible to compare before and after for these particular tests.

5.8.4 Results

	NTFS	**FAT32**	**EXT4**	**HFS+**
Test 2	Pass	Pass	Pass	Pass
Test 3	Pass	Pass	Pass	Pass
Test 4	Pass	Pass	Pass	Pass
Test 5	Pass	Pass	Pass	Pass
Test 6	-	-	-	Pass
Test 7	Pass	Pass	Pass	Pass

Table 16 - Test 8 results.

General findings:

In this test we compared the checksums of the forensic images calculated before and after running Deepthought. All results show that Deepthought does not affect the device being processed.

The MD5 comparisons are listed in Appendix B30.

[14] Logicube WriteProtect Desktop

6 Discussion

6.1 Use of method

The *classification tree method* from ISO29119-4 described in Section 2.4 was a suitable method for deconstructing the functions to be tested. We could have gone into even greater detail when breaking down the requirements for the functions. One example of where we could have been even more specific is defining metadata for each of the listed and extracted items. We chose not to include metadata in the classification tree because it would make the classification tree much larger, with repeating factors.

Using the *all combinations technique* for testing the assertions provided good test coverage for the tests. Using the all combination method did however, increase the complexity of analyzing the results, and could be too comprehensive for testing specific functions at a later stage.

The NIST general test method is a versatile method for structuring the tests. We believe using the techniques listed in ISO29119-4 to replace step 1, 2 and 3 in the NIST general method gives a structured and detailed approach for testing digital forensic tools.

We believe the functions tested in our experiments are important functions for a triaging tool aimed at child exploitation cases. In retrospect, we should have tested one of Deepthought's functions for automatically comparing MD5 checksums of files on the system, to a list of previously identified illegal content provided by the investigators. This function is important, because it could possibly give the investigators a list of illegal content at an early stage of the investigation without requiring manual labeling of the files.

6.2 Datasets and test design

The datasets generated for the tests in this thesis contained a limited number of files. The fewer files in the dataset, the larger the consequences of an error in the tool.

One of the factors to consider in the Daubert standard as described in Chapter 2, is the error rate of the tool. While it would be possible to calculate the error rate in the tests conducted in this thesis, we believe the datasets used are not suitable for demonstrating error rate. The datasets we have used contain few of each file type to act as an all combinations test, to test if the tools' functions meet all specified requirements. Because of the limited number of files of each file type, a single error would have major impacts on the error rate. For this reason, we have not calculated the error rate for the tests. There is certainly a trade-off when designing datasets; A small dataset might reduce the chance of catching errors, but a large dataset makes analyzing the results more complex. We opted for the former in order to test more functions within the time-constraints.

Our method of verifying the email dataset was far from ideal. The test was designed without knowing if Deepthought extracted the files in their original form, or if the emails were parsed, and saved as other formats. Because Deepthought is parsing the emails and splitting them into separate files, we cannot use the MD5s in the reference set to verify the content. Because of this, we had to manually compare each parsed email to the results from AD Lab. This method is not feasible for larger datasets. If we were to redesign this test, we would have made a dataset where we had controlled the content of

each email including the email headers, replacing the content with unique stings which can be automatically compared. Because different digital forensic tools might parse the emails differently, this method might not be re-useable for verifying other tools. More research should be done as to how to efficiently verify parsed data.

In Test 3, Deepthought made partial recoveries of some of the videos. When analyzing the content of these files, it became clear that all files used when designing datasets should be unique, both in checksums as well as visual content. If each file was unique in visual content, it would be possible to look at the partially recovered files and identify which file was the original. Comparing datasets with multiple occurrences of the same files also made it more tedious to verify the data, because comparing MD5s could result in multiple hits in the dataset. This made it more difficult to trace the MD5 back to the original file. When creating the datasets, we reused some of the files in multiple datasets. An example being the *msg* email files from test 2 were reused in test 4. By reusing files instead of introducing new files in each test, it is likely that an eventual error parsing those specific files will carry over to the next test, making the test results less accurate.

When designing the tests, we contemplated adding all the datasets, into a single large test. Although combining the tests would simplify the acquisition phase, we concluded this approach would make analyzing the content more complex, and would arguably reduce the reusability of the datasets.

For future test design, we suggest using datasets with unique files across all datasets. This will make it easier to trace potential partial files back to their source as well as making MD5 comparisons more manageable.

6.3 Result comparison

The process of verifying the output produced by Deepthought turned out to be more time-consuming than estimated. Deepthought does not list the exported files explicitly in the web report, and the files are saved using MD5s as filenames. Because files of the same category are saved in the same folder, without a document describing the source path of the extracted files, it is difficult to trace the files back to their origin.

We have used the AD Lab for generating results to be used for comparison in the tests. Using only one tool for comparison is not ideal because the reference tool might be incorrect, and might not be the best tool for a particular dataset. For comparing results, we also used reference sets made in advance of the experiments. These reference sets only contained MD5 checksums and file paths, and did not list timestamps of the files. Albeit being more time-consuming to create, we believe adding timestamps to the reference sets would be a good improvement when creating datasets for testing digital forensic tools.

We believe using Microsoft Excel to gather the results from Deepthought, reference set and the reference tools made the comparison of data transparent for anyone wanting to review the results. Because there is no standard for how digital forensic tools list output, we think it is important to use a tool where it is easy to normalize data with different structure. This could have been done using a database in combination with scripts comparing the results, but we argue that would be more difficult to review for third parties. Although we believe using Excel made the process more transparent, it was not an ideal tool for the job. The varying output formats from the tools combined with unpredictable forced formatting of cells in Excel made it a time-consuming process to

normalize the data, with some need for manual conversions. One example of where we had to apply manual conversion was when handling the file listing from FAT32 filesystems reported by Deepthought. The timestamp format used when listing information from the FAT32 filesystem does not match the timestamp format used when listing from EXT4 or NTFS filesystems, and was not interpreted as a timestamp by Excel. We realize manual conversions are a source of potential errors, and although we were careful to double check our work, this process may have introduced errors.

6.4 Test results

6.4.1 General findings

Through all the tests, there were some reoccurring errors made by Deepthought.

The user manual (37) specifies that Deepthought should support the HFS+ filesystem. Our tests show that Deepthought is listing the partition table of the HFS+ filesystem, but is unable to present any information about the files stored in the partitions. One of the requirements for digital forensic tools is that errors encountered during the process should be clearly listed for the examiner to review. There were no warnings in the web report listing any errors for any of the tests, despite obvious errors in parsing the HFS+ filesystem and various file formats. We believe adding a feature to report errors to the examiner is paramount for the software's credibility in court.

Deepthought saves all extracted files using their MD5 checksum as the filename. If a file occurs more than once on a file system, Deepthought will only extract one copy of the file. This is likely because all files of the same category are extracted to the same folder, and filenames within a folder must be unique. This probably results in the file with the same name being overwritten, or that the second occurrence of a file is unable to extract to the folder. Although this situation would make it more difficult for the investigators to trace the files back to their origin, we argue that it might still be an acceptable practice because Deepthought is a triaging tool, meant to give a quick overview of the evidence. If an illegal image is found on the system, it might not matter if there are more than one copy of the same file.

In general, Deepthought makes many errors when listing timestamps. We have not been able to find a pattern of the errors, as in some situations the time is off by only a few minutes, while in other cases the full date is incorrect. Timestamps can be of utmost importance in legal cases, because they can greatly affect the presumed timeline of events in the case. An example being if a photo shows the suspect at a difference place than the crime scene at the time of the crime, this can lead to the exoneration of the suspect.

6.4.2 Deleted file recovery

In test 1 and 3, Deepthought was unable to list or recover all deleted files. In test 1 Deepthought did not list any of the deleted text documents on any of the file systems. This could indicate that the function listing deleted files only supports certain file types. In test 3 Deepthought detected and recovered all deleted images and videos in the EXT4 file format. In the FAT32 and NTFS filesystems, only 40% of the deleted images and 10% of the deleted videos were extracted correctly. These numbers are misleading with

regards to the accuracy of the tool, because the dataset contains few files making a large impact of an error rate. For accurate error rates, the tool should be tested using a dataset containing a large number of files. Deepthought was able to extract some of the deleted video files, but with a different MD5 than reported by AD Lab and the reference set. Deepthought recovered these consistently with the same MD5s on all filesystems, indicating it could be something wrong with the implemented method of recovering deleted videos.

6.4.3 Images and videos

During image extraction, the file extensions of all extracted images are changed to *jpg*, even if the original files are not *jpg* images. Calculating the MD5 checksums of the extracted files show that the content of the files is unchanged. The same can be observed for the extracted video files. All extracted videos are changed to *avi* file extension. However, the actual content of the files are unchanged, meaning the integrity of the extracted files are kept. Although we believe it would be better to extract the files using the same file extension, we argue that this is an acceptable practice because Deepthought is a triaging tool, whose main goal is to present the findings to the investigator before the device is further analyzed.

The movies tab in the web report was empty in all tests, indicating that this feature does not work. This is an important feature of Deepthought to list all the tagged videos, making it more difficult to trace the extracted files back to their origin.

The EXIF tab shows the image with the extracted EXIF information next to it. It does however not show name, file path, or MD5 of the image, making it difficult to locate the images in large datasets.

The EXIF information only shows geo-location EXIF data. Information such as make and model of the camera, or timestamps saved in the EXIF information are not listed. This information could be useful in a child exploitation case, because it could be used to establish time of the crime as well as possibly linking other seized devices to the crime. EXIF information from videos are not listed.

Our tests demonstrate that the tool excels at exporting active image and video files, but fails at displaying the correct timestamps associated with the files.

The successful listing and extraction of images and videos is arguably the most important function of a digital forensic tool aimed at child exploitation cases. We believe such tools should be able to recover all deleted images and videos in simple deletion scenarios like the ones used in this thesis. We believe all digital forensic tools should list the findings in a way that makes it easy to trace the files back to their source.

6.4.4 Email

Our tests show that Deepthought does not support *msg* emails as stated in the user manual (37).

Deepthought supports *pst*, *ost* and *eml* files, but does not extract the emails in their original form. The emails are parsed and separated into multiple files per email, and the content is correct. We assume this is a step towards listing the emails in the web report. Deepthought does not list any information in the web report about the exported emails.

The *mbox* email format is extracted in its original form, but is not parsed like the other email formats.

For a triaging tool, we argue that the main goal is to present the findings to the investigators in a manner that makes the artifacts easy to review. This is currently not the situation for the way Deepthought handles emails. We believe the email archives should be saved in their original formats in addition to being parsed. This way the email archives could be imported in other tools directly for further processing and analysis.

6.4.5 Documents

Deepthought extracted all documents correctly from the *NTFS*, *FAT32* and *EXT4* file systems. In the user manual (37), it is stated that Deepthought should extract images from *PDF* documents, but does not say anything about any of the other document formats. In test 6 Deepthought was unable to extract any images from the *docx, xlsx, pptx, ods* or *odt* files, confirming that Deepthought only supports image extraction from the *PDF* file format. Each of the documents in test 6 was embedded with five images of various formats. Deepthought was only able to extract a jpg image from the *PDF* document. When compared to the results from test 2, we see that the document formats supported for image extraction are inconsistent. In test 2 Deepthought extracts images from *ppt* and *doc* files. This could be caused by differences in how the newer *docx, xlsx* and *pptx* formats embed images, as test 6 uses the newer Office formats, and test 2 uses the old Office formats.

6.4.6 Archives

Deepthought consistently extracted only one of 4 encrypted archives larger than 500MB on all three filesystems. Deepthought is calculating the entropy to determine if a file is encrypted (37), and these results might be caused by the entropy of the other encrypted files not meeting Deepthought's threshold. During processing of the archives Deepthought seemingly stopped making progress. This was caused by the program waiting for the user to input a password for one of the encrypted files. What we assume is a bug stopped the program from showing the prompt for password. The encrypted archive requesting a password was not extracted by Deepthought. We argue that all encrypted archives should be extracted in their original form for later processing in other tools, as Deepthought is not designed for password cracking, and the investigators might not know the passwords of the suspect in the initial phase of the investigation where devices are triaged.

6.4.7 Forensic soundness

In test 8 we compared the checksums of the forensic images calculated before and after running Deepthought. This test shows that Deepthought does not affect the device being processed. This means Deepthought passes one of the requirements for digital forensic tools as listed in Chapters 2 and 3.

When inserting USB drives into the computer running the Freetool distribution, the device is mounted as writeable as default. In terms of forensic soundness, we believe that all

devices connected to the Freetool distribution should be mounted as read only to make sure the device is not altered in any way.

In tests 2 and 4 we see that many of the email files are not extracted in their original form, but are parsed and saved as individual files. We believe Deepthought should extract all relevant files in their original format in addition to parsing the data. This would preserve the integrity of the files, and allow for further processing in other digital forensic tools.

Combining the Created, Accessed and Modified timestamps into one criteria makes Deepthought fail almost all metadata tests. We could have decided to separate the timestamps as one test criteria per timestamp, but we argue that showing the correct timestamps is too important for an investigation for it to be acceptable for one of the timestamps to be incorrect. For easier reference for the development team, we have highlighted all timestamp mismatches in our test results provided in Appendix B-1 through Appendix B-29. We believe that the reliability of the evidence would be put into question if the metadata associated with the evidence is incorrect.

6.4.8 Requirements

In this thesis, we have reviewed legal requirements, and requirements specified by the digital forensic community. We identified that digital forensic tools must adhere to the following requirements:

- The tool should not affect the device being investigated.
- The data presented by the tool should have the same meaning as the original data.
- Potential errors encountered during the processing should be clearly stated by the tool.

Our tests have shown that processing digital devices using Deepthought does not affect the device, passing the first requirement.

As argued by Guo et al. (48), a digital forensic tool can have functions failing a verification test without invalidating the whole tool, as long as the functions used to produce the evidence passes the test. Across all the experiments, the majority of timestamps presented by Deepthought do not have the same meaning as the original data. We regard listing timestamps correctly is of critical importance for digital forensic tools. Deepthought fails the second requirement.
Deepthought does not present any information about errors occurring during the execution. This is an important feature for a digital forensic tool to be trustworthy, so the investigators can take the errors into account when examining the evidence. Deepthought fails the third requirement.

7 Conclusion

In this thesis, we have conducted scripted tests based on techniques from the ISO29119-4 combined with the NIST general test methodology. Combining the classification tree method, all combinations test and the NIST general testing method provided a structured method for establishing test criteria and a versatile framework for designing functionality tests for digital forensic tools.

None of the functions tested in this thesis passed all test conditions. The main reason for all tests failing was the misrepresentation of timestamps.

Our tests demonstrate that Deepthought excels at exporting active image, video and document files, but does not support all features as listed in the user manual. Emails, browser history, and extracted videos are not presented in the web report, and no files were listed from the *HFS+* filesystem.

Processing a device using Deepthought does not affect the device. However, the timestamps presented by Deepthought does not have the same meaning as in the original file, and Deepthought does not report errors. These factors lead Deepthought fail two out of three requirements for digital forensic tools.

Our tests have demonstrated that the current version of Deepthought does not meet the requirements for digital forensic tools. Deepthought has great potential as a triaging tool, but in the current version, it is not ready to be used by law enforcement.

8 Future research

As the shortcomings demonstrated in this thesis are fixed, we believe Deepthought should be tested again, using the same datasets. This should be done to verify that the errors have been corrected, and validating the functions for future use by law enforcement.

During our tests we did not calculate the error rate of the functions due to the limited size of the datasets used. In future testing, datasets with a large number of the same file types should be used to calculate the error rates.

Working with this thesis, we found that creating an adequate reference set requires more than just knowing the MD5 and path of each file. When testing digital forensic tools there is a need for reference sets including the metadata of each file. When testing functions that parse the content of the files, there is a need for reference sets containing the content of the files as well. An example of this being email archives, where the content of each individual email should be accounted for in the reference set for comparison. We believe this is achievable by generating the datasets in a controlled environment instead of using service providers to generate data. We suggest research into a framework for creating datasets with controlled content, automatically producing reference sets to accompany the datasets. We also suggest research into methods for efficiently validating parsed data.

We believe it would be useful for the digital forensics community to develop a crowd sourced dataset of files causing problems for digital forensic tools. This dataset could be used when testing open source tools, in-house developed tools, and commercial tools, driving digital forensic tools towards better reliability and error handling.

Through our thesis, we have started the work of identifying test assertions for digital forensic tools aimed at child exploitation cases. This work can be further extended to include more functions, and ultimately ending in a comprehensive list of test assertions which can be used when testing any digital forensic tool functionality. We suggest research into other digital forensic tool functions, to add to the publicly available test assertions.

9 References

1. Caviglione L, Mazurczyk W, Wendzel S. The Future of Digital Forensics: Challenges and the Road Ahead. IEEE Security and Privacy Magazine. 2017;15(6):12-7.

2. Scientific Working Group on Digital Evidence. SWGDE Recommended Guidelines for Validation Testing. Version: 2.0 [Internet]. www.swgde.org: SWDGE; 2014 [updated May 10 2019; cited 2019 May 10]. Available from: https://www.swgde.org/documents/Current Documents/SWGDE Recommended Guidelines for Validation Testing.

3. National institute of Standards and Technology CFTT. General Test Methodology for Computer Forensic Tools [Internet]. www.nist.gov: National Institute of Standards and Technology; 2001 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.nist.gov/document/test-methodology-7doc.

4. National institute of Standards and Technology CFTT. Federated testing project [Internet]. www.nist.gov: National Institute of Standards and Technology; 2019 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/federated-testing.

5. Grajeda C, Breitinger F, Baggili I. Availability of datasets for digital forensics - and what is missing. Digital Investigation. 2017;22:94-105.

6. National Police Chiefs' Counsil. Policing for the Future Written Evidence by the National Police Chiefs' Council [Internet]. www.npcc.police.uk: National Police Chiefs' Counsil; 2017 [updated March 2017; cited 2019 May 20]. Available from: https://www.npcc.police.uk/documents/NPCC Submission Policing For The Future 160217.pdf.

7. Internet Watch Foundation. Why do we exist [Internet]. www.iwf.org.uk: Internet Watch Foundation; 2019 [updated March 15 2019; cited 2019 March 15]. Available from: https://www.iwf.org.uk/what-we-do/why-we-exist.

8. Rogers M, K,, Seigfried-Spellar K, C,. Using Internet Artifacts to Profile a Child Pornography Suspect. Journal of Digital Forensics, Security and Law. 2014;9(1).

9. European Commission - Migration and Home Affairs. FREETOOL Project - Free, Reliable Tools for Investigating Cybercrime [Internet]. https://ec.europa.eu/ 2019 [updated March 27 2019; cited 2019 March 27]. Available from: https://ec.europa.eu/home-affairs/financing/fundings/projects/HOME_2011_ISEC_AG_INT_4000002171_en.

10. Faheem M, Kechadi T, Le-Khac N-A. The State of the Art Forensic Techniques in Mobile Cloud Environment: A Survey, Challenges and Current Trend. International Journal of Digital Crime and Forensics 2015;7(2):1-19.

11. Jones M, G, Winster G, S. Forensics Analysis On Smart Phones Using Mobile Forensics Tools. International Journal of Computational Intelligence Research. 2017;13(8):1859-69.

12. Carrier B. Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers. International Journal of Digital Evidence. 2003;1(4).

13. Collective work of all DFRWS attendees, editor A Road Map for Digital Forensic Research. DFRWS 2001 USA; 2001; Utica, NY: Digital Forensic Research Workshop.

14. Carrier B, Spafford E, H,. Getting Physical with the Digital Investigation Process. International Journal of Digital Evidence 2003;2(2).

15. US-CERT. Computer Forensics [Internet]. www.us-cert.gov: US-CERT; 2008 [updated 2008; cited 2019 May 10]. Available from: https://www.us-cert.gov/sites/default/files/publications/forensics.pdf.

16. Kyei K, Zavarsky P, Lindskog D, Ruhl R. A Review and Comparative Study of Digital Forensic Investigation Models. The 4th International Conference on Digital Forensics & Cyber Crime (ICDF2C 2012); Lafayette, Indiana, USA. Digital Forensics and Cyber Crime: Springer; 2012. p. 314-27.

17. Department of Justice - Office of Justice Programs. Forensic Examination of Digital Evidence: A Guide for Law Enforcement [Internet]. www.ncjrs.gov: U.S. Department of Justice; 2004 [updated April 28 2019; cited 2019 April 28]. Available from: https://www.ncjrs.gov/pdffiles1/nij/199408.pdf.

18. Carroll O, L, Brannon S, K, Song T. Computer Forensics: Digital Forensic Analysis Methodology. United States Attorneys' Bulletin. 2008;56(1).

19. Rogers M, K, , Goldman J, Mislan R, Wedge T, Debrota S. Computer Forensics Field Triage Process Model. Journal of Digital Forensics, Security and Law. 2006;1(2).

20. Association of Cheif Police Officers. ACPO Good Practice Guide for Digital Evidence. www.college.police.uk: College of Policing; 2012.

21. Carrier B. Open Source Digital Forensics Tools- The legal argument. @stake Research Report [Internet]. 2002 [cited 2019 February 13]. Available from: http://www.digital-evidence.org/papers/index.html.

22. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993) www.supreme.justia.com: United States Supreme Court; 1993 [Available from: https://supreme.justia.com/cases/federal/us/509/579/.

23. Mueller C, B, , Laird K, C, , Liesa R. §7.7 Reliability Standard (Daubert, Frye). GWU Law School Public Law Research Paper No 2018-71 [Internet]. 2018 [cited 2019 May 20]. Available from: https://ssrn.com/abstract=3277067 https://dx.doi.org/10.2139/ssrn.3277067.

24. Association of Chief Poliece Officers. About us [Internet]. https://web.archive.org/: WaybackMachine; 2007 [updated March 29 2010; cited 2019 May 20]. www.acpo.polieve.uk/about.html. Available from: https://web.archive.org/web/20100329003056/http://www.acpo.police.uk/about.html.

25. Crime muesum. Edmind Locard [Internet]. www.crimemuseum.org: Crime museum; 2017 [updated May 02 2019; cited 2019 May 02]. Available from: https://www.crimemuseum.org/crime-library/forensic-investigation/edmond-locard/.

26. McKemmish R. What is Forensic Computing? Trends & issues in crime and criminal justice 1999;118.

27. Casey E. What does ''forensically sound'' really mean? Digital investigation. 2007;4(2):49-50.

28. Duke Law EDRM. Duke Law EDRM Glossary Forensically sound procedures [Internet]. https://www.edrm.net/: Duke Law EDRM; 2019 [updated April 24 2019; cited 2019 April 24]. Available from: https://www.edrm.net/glossary/forensically-sound-procedures/.

29. McKemmish R. When is Digital Evidence Forensically Sound? In: Ray I, Shenoi S, editors. IFIP International Conference on Digital Forensics Advances in Digital Forensics IV; Kyoto, Japan: Springer; 2008. p. pp 3-15.

30. National Institute of Standards and Technology. Active File Identification & Deleted File Recovery Tool Specification [Internet]. www.nist.gov: National Institute of Standards and Technology; 2009 [updated March 14 2019; cited 2019 March 14]. Available from: https://www.nist.gov/sites/default/files/documents/2017/05/09/dfr-req-1.1-pd-01.pdf.

31. Teel Technologies. Detego Field Triage [Internet]. www.teeltech.com: Teel Technologies; 2019 [updated May 21 2019; cited 2019 May 21]. Available from: http://www.teeltech.com/mobile-device-forensic-tools/detego-ultimate/detego-field-triage/.

32. AccessData. FTK Imager 4.2 [Internet]. www.accessdata.com: AccessData; 2019 [updated May 17 2019; cited 2019 May 17]. Available from: https://marketing.accessdata.com/ftkimager4.2.0.

33. BlackBag Technologies. Macquisition [Internet]. https://www.blackbagtech.com/: BlackBag Technologies; 2019 [updated May 21 2019; cited 2019 May 21]. Available from: https://www.blackbagtech.com/software-products/macquisition.html.

34. AccessData. AD Lab [Internet]. https://accessdata.com/: AccessData; 2019 [updated April 14 2019; cited 2019 April 14]. Available from: https://accessdata.com/products-services/ad-lab.

35. University College Dublin - Centre for Cybersecurity & Cybercrime Investigation. 2.-UCD-CCI-Cheryl-Baker [Internet]. www.horizon2020.ie2017 [updated May 21 2019; cited 2019 May 21]. Available from: https://www.horizon2020.ie/wp-content/uploads/2017/10/2.-UCD-CCI-Cheryl-Baker.pdf.

36. Toolan F, Genoe R, Browne A, Bergum U, Shaw A. Deepthought: Initial Validation of a Preliminary Analysis Forensic Tool. Cybercrime Forensics, Education and Training (CFET-14); Canterbury, UK 2014.

37. The Deepthought team. Deepthought User Manual version 1 [Internet]. www.ucd.ie: Univercity College Dublin; 2019 [updated March 03 2019; cited 2019 March 03]. Available from: https://freetool.ucd.ie/.

38. The CAINE development team. CAINE Computer Forensics Live Linux Distro [Internet]. www.caine-live.net2018 [updated April 16 2019; cited 2019 April 16]. Available from: https://www.caine-live.net/page8/page8.html.

39. University College Dublin - Centre for Cybersecurity & Cybercrime Investigation. Freetool 2.0 [Internet]. www.ucd.ie: Univercity College Dublin; 2017 [updated 2017; cited 2019 May 04]. Available from: https://www.ucd.ie/cci/projects/current_projects/freetool2.html.

40. National institute of Standards and Technology CFTT. Autopsy Version 4.6.0 - Test Results for String Search Tool [Internet]. www.dhs.gov: Department of Homeland Security; 2018 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.dhs.gov/sites/default/files/publications/Test Report_NIST_String Searching Autopsy v.4.6.0_November 2018.pdf.

41. National institute of Standards and Technology CFReDS. NIST Computer Forensics Reference Data Sets [Internet]. www.cfreds.nist.gov: National Institute of Standards and

Technology; 2019 [updated February 03 2019; cited 2019 February 03]. Available from: http://www.cfreds.nist.gov/.

42. National institute of Standards and Technology CFTT. The Sleuth Kit (TSK) 3.2.2/Autopsy 2.24 - Test Results for Deleted File Recovery and Active File Listing Tool [Internet]. www.dhs.gov: The National Institute of Standards and Technology; 2014 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.dhs.gov/sites/default/files/publications/508_Test Report_The Sleuth Kit 3 2 2 - Autopsy 2 24 Test Report_November 2015_Final.pdf.

43. National institute of Standards and Technology CFTT. Active File Identification & Deleted File Recovery Tool Specification www.nist.gov2009 [Available from: https://www.nist.gov/sites/default/files/documents/2017/05/09/dfr-req-1.1-pd-01.pdf.

44. Marshall Information Security and Digital Evidence. Independent Validation and Verification (IV&V) of EnCase Forensic Edition Law Enforcement and Government Edition Version 5 (update v.5.05d) [Internet]. https://forensics.marshall.edu/MISDE: Marshall Information Security and Digital Evidence; 2006 [updated February 09 2019; cited 2019 February 09]. Available from: http://forensics.marshall.edu/Digital/Pubs-Soft/EncaseFEv505d.pdf.

45. International Organization of Standards. About us [Internet]. www.iso.org: ISO; 2019 [updated May 23 2019; cited 2019 May 23]. Available from: https://www.iso.org/about-us.html.

46. International Organization for Standardization. ISO29119-1-2013 - Consepts and definitions [Internet]. www.iso.org: International Organization for Standardization; 2013 [updated March 15 2019. Available from: https://www.iso.org/standard/45142.html.

47. International Organization for Standardization. ISO29119-4-2015 - Test techniques [Internet]. www.iso.org: International Organization for Standardization; 2015 [updated March 15 2019; cited 2019 March 15]. Available from: https://www.iso.org/contents/data/standard/06/02/60245.html.

48. Guo Y, Slay J, Beckett J. Validation And Verification Of Computer Forensic Software Tools-Searching Function. Digital investigation 6 (2009) 2009:12–22.

49. Pröll R, Bauer B, editors. Toward a Consistent and Strictly Model-Based Interpretation of the ISO/IEC/IEEE 29119 for Early Testing Activities. Proceedings of the 6th International Conference on Model-Driven Engineering and Software Development (MODELSWARD 2018); 2018; Madeira, Portugal: SciTePress.

50. National institute of Standards and Technology CFTT. NIST Computer Forensic Tool Testing Program [Internet]. www.nist.gov: NIST CFTT; 2019 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt.

51. Flandrin F, Buchanan W, J, Macfarlane R, Ramsay B, Smales A. Evaluating Digital Forensic Tools (DFTs). 7th International Conference : Cybercrime Forensics Education & Training; Canterbury 2014.

52. Beckett J, Slay J. Digital Forensics: Validation and Verification in a Dynamic Work Environment. 40th Hawaii International International Conference on Systems Science (HICSS-40 2007); Waikoloa, Big Island, HI, USA: IEEE; 2007.

53. Wilsdon T, Slay J. Digital forensics: Exploring validation, verification & certification. First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05). 2005:48 - 55.

54. Scientific Working Group on Digital Evidence. Who is SWGDE and what is the history? [Internet]. www.swdge.org: Scientific Working Group on Digital Evidence; 2003 [updated May 15 2019; cited 2019 May 15]. Available from: https://www.swgde.org/pdf/2003-01-22 SWGDE History.pdf.

55. Scientific Working Group on Digital Evidence. SWGDE Minimum Requirements for Testing Tools used in Digital and Multimedia Forensics [Internet]. www.swgde.org: SWGDE; 2018 [updated April 14 2019; cited 2019 April 14]. Available from: https://www.swgde.org/documents/Current Documents/SWGDE Minimum Requirements for Testing Tools used in Digital and Multimedia Forensics.

56. National Institute of Standards and Technology CFReDS. DFR Test images [Internet]. www.cfreds.nist.gov: NIST CFReDS; 2017 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.cfreds.nist.gov/dfr-test-images.html.

57. Duke Law EDRM. EDRM File Format Data Set [Internet]. www.edrm.net: Duke Law EDRM; 2019 [updated February 14 2019; cited 2019 February 14]. Available from: https://www.edrm.net/resources/data-sets/edrm-file-format-data-set/.

58. Beckett J. Forensic Computing: A Deterministic Model for Validation and Verification through an Ontological Examination of Forensic Functions and Processes [PhD]. Adelaide, Australia: University of South Australia; 2010.

59. Internet Watch Foundation. Internet Watch Foundation [Internet]. www.iwf.org.uk: IWF; 2019 [updated March 10 2019; cited 2019 March 10]. Available from: https://www.iwf.org.uk/.

60. Bednar P, Katos V. SSD: New Challenges for Digital Forensics. Information Systems: a crossroads for Organization, Management, Accounting and Engineering. 2011:8.

61. Kingston. DataTraveler® Generation 4 - Kingston [Internet]. https://www.kingston.com/: Kingston; 2019 [updated February 14 2019; cited 2019 February 14]. Available from: https://www.kingston.com/datasheets/DTIG4_en.pdf.

62. National Institute of Standards and Technology CFReDS. Forensic Images Used for NIST/CFTT File Carving Test Reports [Internet]. www.nist.gov: National Institute of Standards and Technology; 2017 [updated February 03 2019; cited 2019 February 03]. Available from: https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/computer-forensic-0-0-2.

63. National institute of Standards and Technology CFReDS. setup-july-10-2012 [Internet]. www.cfreds.nist.gov: NIST CFReDS; 2012 [updated February 04 2019; cited 2019 February 04]. Available from: https://www.cfreds.nist.gov/dfr-images/setup-july-10-2012.pdf.

64. Friheim I. Practical use of dual tool verification in computer forensics [Master thesis]. Dublin, Ireland: Univercity College Dublin; 2016.

10 Appendices

Appendix A1. DFR-01 reference set
Appendix A2. EDRM reference set
Appendix A3. Images and videos reference set
Appendix A4. Email reference set
Appendix A5. Archives reference set
Appendix A6. Documents reference set
Appendix A7. Browser history reference set
Appendix B1. DFR01 NTFS Results
Appendix B2. DFR01 FAT Results
Appendix B3. DFR01 EXT Results
Appendix B4. DFR01 HFS Results
Appendix B5. DFR01 EXFAT Results
Appendix B6. EDRM NTFS Results
Appendix B7. EDRM FAT32 Results
Appendix B8. EDRM EXT4 Results
Appendix B9. EDRM HFS Results
Appendix B10. Images and videos NTFS Results
Appendix B11. Images and videos FAT32 Results
Appendix B12. Images and videos EXT4 Results
Appendix B13. Images and videos HFS Results
Appendix B14. Email NTFS Results
Appendix B15. Email FAT32 Results
Appendix B16. Email EXT4 Results
Appendix B17. Email HFS Results
Appendix B18. Archives NTFS Results
Appendix B19. Archives FAT32 Results
Appendix B20. Archives EXT4 Results
Appendix B21. Archives HFS Results
Appendix B22. Documents NTFS Results
Appendix B23. Documents FAT32 Results
Appendix B24. Documents EXT4 Results
Appendix B25. Documents HFS Results
Appendix B26. Browser History NTFS Results
Appendix B27. Browser History FAT32 Results
Appendix B28. Browser History EXT4 Results
Appendix B29. Browser History HFS Results
Appendix B30. Before and after comparison

Appendix A1 – DFR01 reference set

NTFS

Name
Arcturus.txt
Bunda.txt
Castor.txt

FAT

Name
Alcor.TXT
Algol.txt
Bellatrix.txt
Betelgeuse.txt
Canopus.txt
Capella.txt
XALTIR.TXT
XBEID.TXT
XCAPH.TXT

EXT

Name
Algol.txt
Antares.txt
Arcturus.txt
Bellatrix.txt
Bunda.txt
Botein.txt
Castor.txt
Canopus.txt
Chort.txt

HFS

Name
Alcor.TXT
Algol.txt
Altair.txt
Beid.txt
Bellatrix.TXT
Betelgeuse.txt
Canopus.txt
Capella.TXT
Capella.txt
xBellatrix.txt

ExFAT

Name
Alcor.TXT
Betelgeuse.txt
Capella.txt

Appendix A2 – EDRM reference set

MD5	Path
c037e08d9faa00741a4f8d719a651957	ami/POINTSZ.SAM
e8de0c9ca68172048b56194d24f2845b	acs/YESNO.MDB
e5198546b2d5b385ae9ee4df0b5aa489	bdr/PRINT1.OBD
8b350bf4588bbead86263968bed97d3c	bmp/4seafoodc.cdr
2a9a55c4f586a5ae6748d535f2dc7d9f	bmp/911.CDR
b588a8b8695354fd62c2ad7f7bb86cb7	bmp/AQUA-OS2.BMP
46499f952f0de13e7477976c450d181d	bmp/BLUES.BMP
9a0d1ada809b626086cb917e8a0581b0	bmp/C5MENU.CDR
23729576481efe1029a3918ae629af70	bmp/EYE.CDR
2272d1c397707b5d6d2e9f4cf7047e29	bmp/JAZZ.BMP
ebffbc2e89ff705b68f22ca742b979a0	bmp/mdiframe.ico
faa2e0c2853653a849af85d8772c5a6c	bmp/PO.cdr
e23950223a5ff9ec4102eb5d28358187	bmp/SELECT.CUR
7b592b065be39991c95c5292e1356fe1	cdr/5radialg.cdr
f8b6e972a0aac264b28208f2ded6e3cd	cdr/award.cdr
6a5108511b3ecaf85cca12822868826c	cdr/BOXES.CDR
639c7052f802e7729b2e36e8199c27bd	cdr/FROMWPG.CDR
f143c4a13479a9545f2a89624dfe8775	cdr/GOODPEAR.CDR
85bb5d88cff11ab5544b17cd0103dc44	cdr/UNIFILL.CDR
8c96a65ffb4788da4338e3ebc515e153	cdr/itmenu.cdr
fb8709e611aae93f322fc33d38a55117	cgm/APPLE.CGM
8370f08d573ced7c782afd7dadbe32bc	cmx/NEWSFLSH.CMX
3184e900e2642a765809590be683afc5	dbs/BUSINESS.DBF
18c0dceeb707590e8c8348e725494993	dbs/PICTURES.DBF
82fef39b1d00d57f0a65a09e1d3e7f14	dbs/SYSTIMES.DBF
7825127897b82ad6c4d944f6518f69b0	dif/NAVYDIF1
c870181509b049636fe4dc9c833d0307	drw/TUTOR07.DRW
7b0939689b33ce337b3d005785b2488e	dsf/COASTER.DSF
10eb8906c030609347224b968c0316c3	dwg/AIRPLANE.DWG
19bd1d04c14873ba2ca826c2981039e1	dwg/3DPolygon.dwg
a748ecf0d3d49e8e35ee6827618d4c41	dwg/ASHADE.DWG
ccb4a11240da1ca0a5b5b546ad3d844c	dwg/BLKTEST.DWG
e0b01c063a04b32c1d53a06fb90fa1e6	dwg/CHROMA.DWG
345b37583981bf06781cbbea33e4f206	dwg/NOZZLE.DWG
3bf42d0df2f6d9fbe2c112153921432f	dwg/colorwh.dwg
984f724f53627466b095ece0ba2cf9a5	dx/DEMODX41.DX
132460d4ab8ed8ff8629554fe34c9de7	dx/TEST1.DX2
7e5bc5ef74e2442cec50bebeeb671f57	dxf/CHROMA.DXF
7f37ed979e033affc0b7e846d9e892cf	dxf/CHAIRBIN.DXF
0c10d82c8e0854ee2cf46c0b918a7b15	dxf/Icaddwg1.dxf
f3aa3427ff169642b7e7c31d4b7ccbff	dxf/DOME1.DXF
41a41f6eecb4e8d84750ebd593f61087	dxf/NOZZLE.DXF
c39933a0e3ba75c694355fd6a6734c20	emf/BALCONY.EMF

1548b6cebc547649bb809619edeec1d5	emf/ObjectFillStyle.emf
12a66f78ed794b847e4385f7ba0d8794	en4/CHARFORM.WPF
d9ffb12b5940059fe392e93165b4fed6	en4/ENABLE45.WPF
5c9e4b4b1abf61722ff5d0c384041053	ens/FINANCE.SSF
d3c24e69fa187104fb7ab89a531921ba	eshr/dasharrow.rtf
636052555d216df7bb5b8cc94e892979	emf/allshapes.emf
0f66ae5a5205514f19f1f956bcedc153	exe2/MAKEVID.EXE
012b369c58414cc52257e2e71686f115	fax/GROUP3B.FAX
cb250b9c9d7597c473d4c9f356e230e8	exe2/demet.dll
83aefffa7a6ec4eb09eb82b6717b3dab	fcd/FLDTYPE.FOL
8373efda4241f2414956f29eb0cad13c	fcs/FORMATS.SS
4a9b7061440a7b9dc749f4b3bf3a61a0	fft/BOLD1.FFT
1c00df12daced34daf37a4b10b143b1f	flw/FONTS.PRZ
dadaf5ef1970a9cf1cbda2e822fccf95	flw/PATTERNS.PRE
713a2fe0cf7dce3a96c37f3feba8ac1a	fmv/123PIE.FMV
9a8c9ec820771fa62ac84642c3fc9449	fmv/KITCHEN.FMV
d810a4809cf4b468d3a01f64da2bb95f	fwk/REPORT.FW3
d8e92a8c9731a3d611bebf6a1e2b9349	gdf/GDF
0bfaa0aa57de12e2f4356c8c141259fc	gem/TRUMPET.GEM
d5861aa2c3cab4e0e0a59818fc93b397	gif/A_BITMAP.GIF
199d62574a7220a81e687588717bc723	gif/bannerimage[1].gif
2423fb436c266c7cbf03d42b6d275cac	GP4/SCALE.cal
ccb1929bcf9296932f0926f0fd1e61de	gzip/uedit32.ini.gz
d655c216fdf41d2c335a96bf3eddaa8c	hgs/ORG2.CH3
46d0c5fe8dd36c0f1ecc042a38188740	hgw/OBJS2.PRS
9976931bbe1ddb549a10a50f271c6976	html/GGGR.HTML
eae2331012ba3af7aed873171cac52a7	hwp/Korean_tab.hwp
fd230e8ee10d319aa8fe4247eb050964	hwp/LESSON2.HWP
515d78ca022824142cd61635f865dd7c	hwp2/TextBoxes1.hwp
bb37db14d31dfa0761ba45bc8fa1c081	fpx/catbird.fpx
cf08bee94b8e123bb7f0c01d8136aef5	ich2004/flag.JTT
84eea30330339b5d189ae9c372dfeff9	ich2004/header.jtd
402d36b11506e418d5902f3746676b6d	ich2004/overlay.jtd
798bf01407741fd4ae818f906ef316b2	ich2004/miyazawaKenji.jtt
718d42260591565892f1c422528050ab	ich2004/welcome.JTD
c41902277f624a4607dfa310cbfab01c	ich6/NUMBER.JBW
6f67adda035b40c6ceadde84f9c1ab53	igs/F408X.IGS
0d4759a897a30b876559e6b4a8b4b7dc	igs/IGESD.IGS
e387ae23d1cd01d0219a840f329a1a54	img/RAZORBK.IMG
0e8062eca21562437094f59cefea0104	iwp/ASCII.DOC
04e7c4066325d624951fcb1b421a4766	iwp/CHARFORM.DOC
38d8f65cd9bb1a85d740fbe51dbd183b	iwp/INDENTS.DOC
822ae7af34f0de1866f25ca4a8f9698c	iwp/MANUSC.DOC
561828f83519876fd4f37eeb4535ed22	iwp/OUTLINE.DOC
999b15165231fc37dd963eb58784d661	jbg2/003288.jb2.pdf

63a3c7d05f61b7efada843a86e3003aa	jpg/leaf_BG.JPG
6eac817feff5a8bf288f2605381eb5b0	jw/COLS3.JW
d644c3358a37669fadb06575e94160b9	jw/JWRITE2.JW
ee708a3ac23273c2aba3c1d0fbf8ae1b	jw/DEMOQA3.QW
27d72d7c7ec79ab1442db82de05d0e76	l123r9/Place Inside Plot Area.123
1a33e54639902cfddae2edd2a7866ea6	leg/GGGR.ws5
a39473df4757643c41967fb2ab042262	leg/LEGACY.CHP
ede2114aa760e289ffa9efe4773d1f66	leg/GUIDE.WSD
ec951c4c69b2f38f68d21163311bf52d	lwp/complex.lwp
b6c85296673ac4f0a64adb6d53b52aca	lzh/H.EXE
2e2bb75fb3ef99689bf88e858a9099cf	lzh/S.COM
df4df3765181c2911af7c6d08274fa15	m11/TEST.AF0
15ad4473dfa891c0035d3c4749bca7cd	lzh/WINSOCK.LZH
3e52b307e39b890af8090b4279c2fcf6	manu/SAMPLE7.DOC
3307d3c4a31dd85238fdc289ac14af9f	mcw/BOLD
df47a310fcf7be33429302902e4fd3e3	met/PAPIE.MET
1fddf61a05af40ff8f83f71c6a601943	met/AIRPLANE.MET
b38789a0c48ad78d6629623896daf7f6	met/USA2.MET
6b9c6f5b0dc866efb826e1f9e2e213a3	mm/COMTAB.DOC
60112b51c3ed5919a456769aff7dd0c4	mm/UNDER1.DOC
bd790257ce66e877833e383aecde17e6	mif/46226.MIF
f2c166c0f2decbc2ac556b1d71c5490e	mif/MSBUG2.MIF
501cca369ce2c2824184e940734e5b88	mm4/MATH.DOX
e0cf58e2222eac644b56134c5b42a21e	mif/TAG2.MIF
b0a2fd102b79dadb595d31b4529a3343	MMFN/BLD2.FNT
584d47aacd355fa0a202587c33aaa664	mp/MPLAN4.MOD
4119e44a57c2493fb9a640aa1e6ff464	mpp/1task3resources.mpp
54e1d61ff684923292d60261c4344638	mpp/DrawingWithoutTasks.mpp
0852bda4a6eaf5c78d9999470603ccb3	mpp/ProjectInfoWithBaseline.mpp
de3a19cab697c2f7094a5b38182796d7	msg/ManyFields2.msg
f49c27b05d5d355442e457f339510294	mpp/ENGINEER.mpp
d4dc90d5a2f8f5eca6ee5a879839e4d1	msg/ManyFields3.msg
bfbee2f23b7a725c87f9e3fdc333414e	msg/ManyFields4.msg
7ba9662b5a383f8472f0230312fc5f6a	mif/process.mif
388e44b47eeb174e9c9b28390facc926	msg/ManyFields5.msg
f001c50dc4f6c798bcdc062d6d3fbf89	msg/Pink Note.msg
d584b2114c64dd140e784bdd43485360	msg/Regular.msg
214fdc27fac4506680b7296c655e4bc2	msw/CENTER.DOC
811ac4dcf21aab1c3db9852f53b36a04	msw/NARROW.WRI
a323465278eca02d3eb09b295064d0db	msw/TABLE1.DOC
1d650dbc8caebcdfcb158f59bd71881d	msw/WPRD55.DOC
2517d060c1c53f436ab46b43addad742	mwkd/MACWRK2.DB
f192593ebd4b4a3c99379d27b0576f16	mwp2/TABLE2
98099682a1c95a11e157fa2a8d39f666	mwp2/WPM31_C
529c04c8d15cf9e7421b3d80e132645a	mwpf/WPFFORM4

4e2b0926fbf96a01cee8e1de7bd1b338	mwrk/FLRIGHT.MWK
380fb42da4339117a97ac66ff27e76d6	ow/OWTUTOR.WP
bae5109911821269552cda40af7f756b	pbm/disk.pbm
053478c1eb71707de6560a43fa37cc86	pbm/A6.ppm
961916be7b5fb25de7ac4b6c35378b5e	pbm/gnu.pbm
f287dc8e7aad22d28a2b8365d009b5bc	pbm/FL.pgm
9b35f41bb32cfc2634721d02ac1b74b6	pbm/mab.pbm
0b43d0815d0238f21ae99df0a6decb22	pbm/tmp.pbm
080a065057c4775cd6eaf101d7b004b2	pbm/Ie.ppm
f823f0ba5f631aeb96c03aae091e3ae4	pcl/PCFILE50.LTR
3dee5d5809708b25899a1a3dbe871454	pbm/owl.pgm
b754692f45816c25dfe3a2977ffd35e5	pcx/WATERFAL.PCX
977402a527c66cb64958fb40cc01697f	pcx/SKELET4.DCX
4180218ce53e62923ffec47598c1d802	pcx/APPLE.PCX
bbd1dd45eda584945eec86bd9cc81d7b	pdx/ATTFILES.DB
99aff456c33acfeb2cbd76c06145b625	pdx/BOOKORD.DB
0b72ec9e79c7a8a93d9cd318a2ff0b84	pdf/LEAR.PDF
be25eaeaa5bf6265e4ae7bee7ce582e7	pdx/NEWRECS.DB
846f07416444eba27e70c5f59ca204ef	pdx/SYMBOLS.DB
b49394ce6e3f709deae7e557b00a2f5e	pdx/TASKLIST.DB
b73b60c7c0b43f52a49cbea3e76e9fb4	pdx/TESTPDX4.DB
780ee10ee4f178ce18158c604b4aa52a	pbm/d1.pgm
966030a00fed735660a07e5094ee696b	pdx/TESTUM7.DB3
d6d479ae1f356cf4f6d9058142b1fc26	pcd/ICESTORM.PCD
b8e1a78da00a35c453064ceff0cfec66	pfs/DEMCHOIC.DOC
34962f6054bf6ce2e5b40831f5e25751	pfs/ITALIC.DOC
2b2d13f4467ca6565e32760e6a046ea8	pfs/MANUSCRT.DOC
d1be2ff24dccd0d806ae2486621f2692	pfs/SAMPLE.IWA
5cda31a281996f8ef282c98e65236c4a	pfs/TABS.DOC
f44b4428a1f3b21c88c05ac67b935492	pdf/artistic2.ai
0f5295fa028301fca19923ce9f188fe2	PFSWrite1/GGGR.pfs
3d79151aa6e654d889e0d1eeaf2516dc	PFSWrite2/GGGR.pfs
3e4b2e3637aa5ad7269e73d29cd641f7	pgl/AN2.PGL
d1541e9786fbd903963e7c1ea964f9ba	pgl/CRTOON01.PGL
5ee2a1ed2b41a540aff31dda9d82276f	pic/3D.PIC
20bd39cb75bb215c0fd0c699ceafe034	pic/BAR10.PIC
06cb3c27cb661d47689c3b4d58dfdb6a	pict/DOODLE.PCT
64497e144ca9253d8e169a1eb9fce30a	pict/BALL-16M.PCT
f85682be6410cf32ccd4ae27e985584e	pif/BALLS2.PIF
55b93d8d7915329d2c7206b6b19b5481	pif/PAPCOLOR.PIF
59c2a1e0a58659ada45bfb34e1f47af6	png/pinata.png
c267d936766e338db052cd9b0bfbf1e3	pntg/PIE.MAC
c0791233870f6eeb874ee7f0ffcf7634	pp2/BLUEBOXC.PPT
0eacd438897c7983e332e113595f4f84	pp2/0191-385.PPT
6a8309eb1511f2c0960dd28a0f9f55da	pp2/OBJS3.PPT

42ecc8d72b67aece04f9c2318cc8e14b	pp7/POINT12.PPT
bcfe706b51b0b00ddf303264165006b1	pp97/Background.ppt
68b47baabfab01f399b1228f3a4d1083	pp97/Creativity.ppt
29bac50046c1fe68b6d20103cc11900f	pp97/Properties.ppt
be6eb35d49713a711926ffdf4ec566d8	pp97/AutoShapes.ppt
b4ccc9b6ce8f16b53c5537334004cb45	pp97/Presentation1.ppt
a90a622e3b22844374264bd2255786a7	pbm/543x543.ppm
238a47e33eb14694bce9b74cf829aecf	pp97/titleandchart.ppt
3ffdf3e026465e61eb62411664c1ae84	ppl/COLFMT.PL
47bd0f1de22c096b9e5fc893517893e3	ps/3DIMENSN.AI
c9b6cfb74ea2842b09f81b13402dc092	ps/SPIRO2.PS
04e84a978cfeffcb13ad86f915941082	ps/MASK.AI
f96800fbf0f8f43c5a63a29c4b5188b9	ps/LAYERS.AI
00592a3823aa91f1bcfaa8e8cf569234	pst/sample.pst
9c00aa74c7c6477c6a6beb1c8f3a6339	psd/85x11.PDD
0a3fd73ef26cf1a0019876dd1f75956a	qa/TEST8
521be360539ec56ecbff3a13c648f5bc	qad/CENTERED.DTF
b96abf1ece324cfe3d3edd6b1950b164	qp6/CHART4.WB2
2ad524284554c51ebf8a92aff9bf0f86	qp6/LESSON5.WQ2
27adfabdcddec15e92b10a32859b1fad	qp6/cashbudget.wb3
b8bb57b780a8e26e163cdcf956fa93a3	qp9/fontsizeand.qpw
fa1a95fa4b9c9d5503a54833925f71fa	qp9/insertcellname.qpw
7120bfa834fa5c6b794164d15bd98d07	qp9/insertcolrowsheet.qpw
84200340aa2a930d5e68006f34f79ad2	qp9/qp9SPEEDF.qpw
88165a9f4016e11fd9a27f733a4ef984	qp9/wines.qpw
4e791918ed28330fdc6a002e044db846	psp/Beach Scene.psp
08cf0fd52fac4888bbf5fae1cfbc6dd2	ras/AutumnBear.ras
0505c0df790006af9c686555acb6a266	ras/BIRDS.RAS
afbb739a109fe172ab478b184f75e50a	ras/PumpkinBear.ras
251c5883485c9833cdef0c1faaf4f524	rbs/MAXCOL2.RBF
4ec284022ac9b646d6f4e83dbc2d5503	rbs/MINCOL2.RBF
f887bec0fe96e94724ca3afcc321de77	rft/INDENT.RFT
7a62e481a6dddd5d60e24cee74048aab	rft/SOFT1.RFT
8b9b68e3b7f04bedd9fac402c93d1258	psp/Spider Spin.psp
9a4299426c222b4a0624bdd66ed0c857	rfx/BANK2.R2D
f67bd311cbc21f325cfb47a2ab860396	rfx/CHARTEST.R2D
6fc89b022055869be56ca919faef1f4d	psp/Mountain Lake.psp
6e0b3243a9c661299263a4974fc85a17	rnd/KITCHEN.RND
e56da54caa22427eecf77f7c3710fe7d	rnd/INTER.RND
c1bba2490704fc9b893f009d9f629d42	rnd/Pushpins.rnd
58fbd1a503863c23557479ce61ba5f4e	rnd/Robot.rnd
5bcf974a1152546cda39bbb1af8919cf	rtf/MSRTF.RTF
7295bc78ae13c9496b63ea48142bd305	sam/DPCT.SAM
0891e47a64a5704fbdcaa40dfe27e183	sc5/SC5.CAL
b689819e2feddb4e927afaaa1d1b4605	sdw/goldstar.sdw

d26cc260a1ab66dd5cafdba023513297	sdw/moon.sdw
42f8738ade9044fb6fe4f1d0f620425a	shw3/arrow_shapes.shw
83faa0f674a1c6a7d781c765ed2b0abb	shw3/line_spacing.shw
889a6312c21063d8bec2d38f0bdc8eac	shw3/lotsofsides.SHW
32d550f14ae51ec3b79ed9aca04c35d5	smd/SMARTDB.DB
fa09513246a1fd6048f3f22bb73033db	sms/SMARTSS.WS
394c3d0011ca3a1629806ecfae950f6d	smt/FNOTE1.DOC
018e8430ce293b90947afbfe730d0679	pp97/Sample_presentation.ppt
3f18ddf66beedd03e5876f0209b32de4	snap/LOTUSNAP.001
facdf29953ef249624f0dbd5374f682e	soc/DatabaseFunctions.sdc
b9a0ed5ffb82864b74ba87ab29dae73b	shw3/A day at the Zoo.SHW
ab1f68e484684a7cef4bbc0d0032863c	soi/DisplayQualityB&W.sdd
f61e215f256dfec8adbc51b9e455cda8	shw3/simple.shw
ace6e0b05072b968c464b34e6c5edf76	sow/footendnotes.sdw
115cbc883603886fceacc7d8da5114d4	spt/SPRINT.SPT
6969c412c4e7093dd3c3efea5f9c1405	taz/1000FILE.TAZ
c601184e96bafb99108ec20a1f5f23f0	text/text.txt
a71dc0355030e8d5e57fca832d5795a5	tga/CHEETA.TGA
2654e501286416df06a66ec04ddbf958	tif6/AQUA.TIF
2fc73af6c5fcc70fb07c792bd4549f3f	tif6/BIKE.EPS
44ed81b15a2f00801366bbf7bb0ad56e	tif6/BROOK.EPS
e3dd8c7f8270e1f414bb3dfaff118014	tif6/CITYSCPE.EPS
2bb02f07a555d3d924f3a5a8578c4d15	tif6/BALL.TIF
8182969c65c6f946b042d3f9adc1d274	tif6/EAGLE256.TIF
9d02d2694a0267444a07d1e3c718ba29	tga/MTRCYCLE.TGA
f86aa9dd4992a146471612cc54775ca0	tw/ED.TW
22faec5fdccaf7332588405230358b2f	txt/ADJLNEND.TXT
0db61b8012f79b8edde85b0c9e0e7b77	txt/COLUMN.TXT
a9b54490720943da9c64267173a7f548	txt/DEMODW4.DOC
8d208e289a1e430017b62ddc866588ff	txt/DW5.DOC
1a33e54639902cfddae2edd2a7866ea6	txt/GGGR.txt
fd7c111a9e2bdd519ce42cd51d084d30	vcrd/CCD development meeting.cal
02718e91639ae1bc0aa316d7c046b085	taz/1000FILE.TAR
44810a4dd532ddb285d049034684dd9c	vcrd/Newt_Gingrich.vcf
ac46124d15bbc421ed508f639bfa1e58	tga/EARTH.TGA
2be96291bc90b5a5cc123f66cd03fbb4	viso/2shape.vsd
2d49f0e8a74c60afad26dd02142a0fea	viso/50 US Existing Access Replication Structure.vsd
358f330c93dbe163c181b79638b8ee08	viso/aircond.vsd
d162ec71b668a3063598aeab6205d86d	viso/archscale.vsd
4104734fceb28c7514ba70b89d32951c	viso/arrows3.vsd
1641c6e438e258278db49fa3d39b2ce1	viso/axis.vsd
21755ce45f620d371b36ac1fb4350386	viso/BlkDia.vsd
12857e82d087bb8a4ef43cd3952086ad	viso/draw.vsd
6ced4c9167cb5e785ccc0900aa7d6927	viso/Dataflow Diagram.VSD
c79c0333caea13a09c0f3f86be186b4a	viso/fontchng.vsd

81bdc683508083697f2751dc51f093ed	viso/Fishbone Diagram.VSD
3ecfb810d20fa2d260d0d6df7e2308de	viso/Jacobson Model.VSD
c62f00774032cc3efc432ff46ac84202	viso/Shadow-Backcolour.vsd
b579e47267befaec0adf8831d467fb3b	viso/starts2.vsd
1d480d193e1d9b98b763732d3735fa61	viso2000/gradient.vsd
9f69515fbae1e2c27b549d0878545c2a	viso2000/city.vsd
4d626b9d0ba4842297a1871a4203593e	viso2002/linecust.vsd
747e69693ac4cf54755d81552837607c	viso2002/3Dshape.vsd
ecf2dbf57f4c0b3cb6c9230a9a64ef61	viso2003/04 Cross Diagonal.vsd
dae2c53fc135841928dd0e643dffc050	vw3/01560.VW4
ac1918eb83adcfca00238556f9150367	w2004Mac/Flyer
c492e8bbc87900318576f058e8bdaa35	w6/ALIGN.DOC
b5f6355c028ce9bb26aa0b1dc2428940	w6/GRPHBORD.DOC
3a811f935a2bab85127d1c0d3fffcf3a	w2004Mac/Letterhead
c9f580db58f567d9757293e2ba307002	w2004Mac/Newsletters.doc
5193b747be332cbd92e5624b6f16fcd2	w97/CalendarWizard.doc
988ee22e3c96ad391b3d4827ca837b44	viso2003/OrgChart.vsd
81dc49cb61770cb64f15c1d0607e4609	w97/Columns.doc
19da282fdfb3be47d49f3b00931b4df3	w97/HANGING.DOC
4574600d3b546ca15286567019d42829	w97/headfoot.doc
6aa61bd20d1d93afa591af86b6abb224	w97/J_Word97_6.DOC
3b2fde2a9fc3cfdb9f079ca281e8c410	w97/InsertDiagram2.doc
ccb713ff3425eb134b6f6d086b4f6e20	w97/PNUMBER.DOC
16ae7c300b4477b1aa1a483fa4ccc8f6	w97/ProfessionalReport.doc
aacf37e664f03f34b1a920d134fd01bb	wbmp/Koala52.wbmp
5ea54d5846867af01f97fbcfc2cc1e39	wg2/123OS22.WG2
aec452d75b5be3089578302bbe94aa3d	wk3/TIME.WK3
e316ca6c1e684a7792cf9c20e9b00cc5	wk6/Seasons.123
900e005e9c319e84a98763d203d9a266	wks/BALLET.WKS
8f67cca816d2072da531247df0d88e8d	wks/CHARS.WKS
48a8396178ee6111adb1ecee27fa26e6	wks/COMMON.wks
79d93ef42409d6ed5b6a2a6d34dcd75f	wks/DISTRICT.WKQ
164baa7cbb1803d80efb89d8312b5f8a	wks/DOLLAR.SYM
830d7dcce5a2ca2c911507d76dc9c39e	wks/INVDB.wk1
9dbc7976b1f5a5df1f77dfd77dd4f5ef	wks/INVDB.wks
b06250c720fd5fa13466787b14edba0f	wks/INVENT.WDB
583de9a6003d8411c5b43180dc4c5fd6	wks/LABEL.WDB
100faaac09d254e54bf8deff4c01e1bc	wks/ORDERS.WKS
3d070c332db4aade2e465d0f1ce0dc55	wks/PRODUCT.WKS
9350dc021113de9f9f2248bf591b8ca3	wks/ROTATE.WR1
a4ea4aa541abe719606e34b2181d8fdb	wks/SAMPLE.WR1
64a9233ce4e0d2753e37fc9f56d019bc	wks/SMALL.WKS
9350dc021113de9f9f2248bf591b8ca3	wks/SYMPHONY.WR1
8778ac56e1396b5b73a971188780f0e0	wks/WRK95SS.WKS
0a2dc016aafdc4854d27e416a4837f82	wm/DEMOWMC.ORG

fe92d462e317820fbaae195358201ea5	wmf/AMHAPPY.WMF
37f65586ba11422600ea3d963f783934	wmf/BIRD.WMF
919b0b7b117a7b7e31dbf962c825e182	wmf/ASPDEBUG.WMF
ca218308957304188bffc751fe7656b7	wmf/COFFEE.WMF
01ab484178de553db8fbee6a134ff5c3	wmf/DEMOPWP.G01
cfc1af521b16503e740300b3362a1931	wmf/preview.wmf
69a534c977ce1dd5c45077dcc083cb8e	wml/Logo.gif
45efc24b01113857d890354f17757a1f	wml/gmeyertestalltags.wml
69a534c977ce1dd5c45077dcc083cb8e	wml/Stellent_logo.gif
199414e654e7da039172eee6606ec4e4	word/AFTER.DOC
817fc791f62b589901ab81b90134f97d	word/ANTATION.DOC
eb32c795437ec2ae8d1456aedadc527d	word/COLORS
b9f5369bdaa4abce80265ed7f7e309b2	word/COMP11.CLX
dd7ad7b821a178dacb3993b037198fa3	word/Macword5
ce8090e0e6888a2add3ac82c78ffaa45	word/PAPERTIT.CLX
ae4b366d381ce05979a1a899f32bbeff	word/STYLE
a3202daec3d22b4414da56a51057f0c5	word/WRDCHARS
57d1a821986b6da322593bf7801c9c55	work/BALOU_C.WPS
e8930d09ae1045a16d79fb787559d298	work/FORMLETR.WPS
433802af1dd9b787003e4bdabee2c4f3	work/INDENTS.WPS
2bc3bc5e93da7dae12959c4bce7719db	work/ALDEMO.WPS
d3fd0caea04de8f93920d5ab81e8fcfc	work/LETTER.WPS
cb5d0e0f27128cd0ff895fad32e7f43e	wp10/datetime.wpd
4133b40c1385e69da2dbe9341caacdeb	wp10/Naming.wpd
1d062b226a16f5b7bf1c354d14a8bbe2	wp11/BasicTabs.wpd
524ffd899dc71b9cf9c16cf0ea29f0b7	wp11/insertGraphic.wpd
06c29f4a75be04400c24faa9f0649744	wp12/All Font Sizes.wpd
2c8097f3aa68f15c637f57cd97ed4a74	wp12/Colors.wpd
ca25196edfe93526aada68218145814b	wp5/EGE00B3.WP5
bf85a110bc85faaec225cd3a88a611cf	wp7/Text.wpd
8df590b76bb6d5daebac4eb11c5ba7dd	wp7/CWP7FNT1.WPD
8870b6db31d3792dc904ff7a980a5190	wpf/INDENTS.WPF
2eb315951e81eb8bf23d57a65b625602	wpg/ARROW-22.WPG
4de56e51f0665b7a395abbe45973b791	wpg/BICYCLE.WPG
2a068de32adb6d656e1587d9c4564223	wpg/NEWS.WPG
513c9b47e60b79ce1a76ea764d00eded	wpg/THANKS.WPG
11c8ce9f74bf97a9c3b7a87196c5b905	wpg2/COT1.WPG
2650d6a87a99fc8d9b90eebaf7ffb459	wpg2/FACTORY.WPG
790435da34d27e0727e444ef734a84e1	wpg2/Gradient.wpg
f4ace7bf5a203776412ba2dce81e6c66	wpg2/TREE.WPG
ff962071f2de4a166e001596becc9b38	wpl/ONEPAGE.WPL
162b29cf8ef9029a991058ab27b81855	wpw/BORDERS3.WPW
9230616122c32d0c75b807e8491a3a76	wpw/FILL2.WPW
1d26e911b34c786792091eebf2dc131d	wpw/TEST.WPW
fd5a80df73368da9877377e5f5d8d85a	ws/11235.BUG

2b6c21cb0f986837bbfa880ef2329dd7	ws/ASCII.WS6
cf29192ac6b18d9bbe6090a43a49aa99	ws/BOTMAR.WS
1e9a840059b5044b41d9b450d25027fa	ws/COLS.WS
1abf0ff6bc28e53449cd63a64cab4200	ws2/WS2TEST1.WS2
9fbe3b7692aa515ccc9b6f00ef17a24b	xbm/TEST1.XBM
f61e215f256dfec8adbc51b9e455cda8	wpg2/simple.shw
3a6140c9b8bbecf6455712b06287711e	xl2000/Bar.xls
0e713ed655141c8de84da6b73cfa9b0b	xl2002/linecharts.xls
fb7390e701058c9d7ea99d4f9ce49f26	xl2003/column.xls
34354e74ce8d1430dda30e4387fc649a	xl2004Mac/Investment Calculator1.xls
e41892052fc081004d396aa598b5e9f6	xl5/CHART64.XLS
d008bad3036d505d65d0528918865e98	xl5/CHART67.XLS
f12b57d1c6597dedab300ccc0009aaa4	xl5/MACXL4.XLS
01249401f5524853a4685356268d1bde	xl5/MANY.XLS
10c75bf5068a776e650767fa30099247	xl5/mergedataerase2.xls
0b031565405c8697e4a5d6b4e911fb47	xl5/XL3.XLS
ba5741c44a31c52c89f6bd9880f307ec	xpm/35FLOPPY.XPM
6e92a86f0638191c9f87272fe57e4782	xwd/SMALL.XWD
dfbdb30680466f6aa21c96bb75e05d9c	xy/XYMISC
6b37fc8470883ae489f5a365b4094e70	xl5/XLS4.XLS
a9bcb32023f84dacf77127c97404b00b	zip/PCT.ZIP
fccf915c044269617a1e832000074863	zip/testzip.zip
3b285ba4c23a4cf691c2e5fa4e82cd66	xl2004Mac/Lists

Appendix A3 - Images and videos reference set

MD5	File
5666765e39fb1c25546ed97101edc481	image files/amalfi.bmp
1c80229fed93f8dde8f520bc5bd4dcfc	image files/bamboo-clump-anamated.gif
9ced30a852058688ca795190a376604d	image files/barn.gif
eb6922ce4023fdda4b2bb48a43e9ee21	image files/cactus.png
95d1f19acd4a1ec7940873ebb0109ba7	image files/cutty-sark.JPG
bf73717421515fb5bf5a63a30c577e29	image files/delete-blini.gif
0525e025d30d3fd01e35adf049f8a1f6	image files/delete-boudicca.bmp
a1bbcf7404b6d68c20acaf2f91d8ff29	image files/delete-cave.png
5bcd00c6f6f98bda479f12f77c8196f5	image files/delete-iris-white.tiff
af64fa519b9fce6ec50ed4e01a825f0f	image files/delete-jump.jpg
9114783f06cf23b47e7b5f2979f8a5c1	image files/delete-lavender.png
ea6f8a193f00736a5f7e4e8a5fba83c9	image files/delete-oak-snow.jpg
2f9930659586b9d3ba0e88ebc7a34547	image files/delete-slices.tiff
bcc35f2a47805b72b1668244e48d4dc9	image files/delete-tapas.gif
d2124379e4be906103560ffca9953df3	image files/delete-zen.bmp
9216e4573096d96ce86bddf1a75d71a0	image files/eggs.gif
7278fd6d4caca870d24d8d77523695f4	image files/forsythia.png
a3cd23180ec6231c25247c6759398e37	image files/gallop.tiff
74edd1e59f69503f6ceff2dc0d11f4a6	image files/injera.gif
3176448ba43cdc76c1e425a24ae4433d	image files/iris-lavender.bmp
e461d275762f718f867fd2327b3dd745	image files/iris-yellow.bmp
2409d6c15dda5cd7c2696a02971ad0dd	image files/jack-o-lantern.tiff
65a875be64f4cc69838c56b5656066cc	image files/lamp-with-geotag.jpg
9a7aabe7fbe1de00cef911d5737a4181	image files/leaf.jpg
20a1bec8e0e36193fe61594434323616	image files/log.png
d52146d20e71e107d7f6de9c375e0b01	image files/orchid.png
69f3c33bcf250835cb8b701374b29388	image files/piazza-dei-miracoli.JPG
82f075b1051e072b49e10df03891bcb7	image files/pink-rose.bmp
f425dabfb0dc6c294f4658f054f7cc21	image files/pisa.JPG
3022e920030df9be408eb36b11ebce05	image files/river.png
d47961e5390c9df9357cf5bec7994fff	image files/shoot.bmp
e575d243a40f16ddb19238623155c055	image files/sliced-tomatoes.tiff
5a0182fe2ef547ea3eb3f0d4d318c552	image files/smoked-chicken.bmp
bc417bbf640cb1f7e710625d5e809236	image files/SPQR.JPG
16a7a75fd4c6af0ab5a8163b204a2add	image files/stonehenge.JPG
2bf4150e7c2008ea17d59e1b472486f4	image files/tomatoes.gif
d0aa8409fc4a0fc1a2385aa8cba02007	image files/trail.png
db28afbfe75f5db14072271fefea035b	image files/tulip-red.tiff
dcc13028116e3f7b986d23525095d4ae	image files/wat.gif
f5917ce264b0c750c93430cd0f3b0e07	image files/winter-street.tiff
439494cf2c5891019dcf73e982d1b73d	image files/yuck.tiff
62dca9c26e2424bf9d60fcfcf978cc0b	video files/20190309_125906.mp4
20c51faee08ec30774cd2f0bbe048f38	video files/bamboo-in-breeze.MOV

bc8149965433ccafe8952686ead831df	video files/delete_fws-video4.avi
9b7a1756e386dd0112d95a86bd45b26f	video files/delete-forsythia.MOV
e7497d810bb02ae6ed2e095623fe889b	video files/delete-fws-video2.avi
969fce07b25caa408c4173c3e3242eb3	video files/delete-more-seedlings.MOV
77d5120b83f3e42f7da6ada9c226d41f	video files/delete-nasa_video3.ogv
e4e5016874138ee98bb485ca3037e5c5	video files/delete-nasa_video5.ogv
5564c9cd8e0c15590726b78c83df4126	video files/delete-noaa-video3.3gp
2d183b34ee9707c39fb37ad78bdcca01	video files/delete-noaa-video6.3gp
64efc1ca23b3254de72c676690153d33	video files/delete-nws-video2.wmv
f2655d340d663714ae023824eeeadb55	video files/delete-nws-video5.wmv
4eedf259896db686dc85e06ead7fef4d	video files/delete-USGS-video2.mp4
157a6983de014c6de47fd3e2504e2b8b	video files/delete-USGS-video5.mp4
4ea9712f8cbee5ab8f26f0f2ba6d1c97	video files/fws_video5.avi
985cb2659775bba546deeeae35aa6539	video files/fws-iphone_video.MOV
704da35f2a12ee9bb6d77c3afcc6afb6	video files/fws-video1.avi
1cb0769184113be1f36c607c6fa299c4	video files/fws-video3.avi
89d18c58de7e9235e2f67d5283ac5ced	video files/fws-video6.avi
e310d0c4a7154cd3f451511f68c42136	video files/nasa_video1.ogv
ef4e8ea02957d2a77c85a6521e07b22a	video files/nasa_video2.ogv
a7f182595666d67dea948b8ab18824fe	video files/nasa_video4.ogv
e03f4c25b4ae4e9972932b03298b8a72	video files/nasa_video6.ogv
f90f29c9d9501dd174b42e1425389ef0	video files/noaa-video1.3gp
1d0916e93409a45101483ac8d9c1528d	video files/noaa-video2.3gp
81205c97c066a9c1178e1f9ea33f65b7	video files/noaa-video4.3gp
80264b4ae9d8460e1d3c37123ed78958	video files/noaa-video5.3gp
4faa10f3150d6102b05a45c9dd35798b	video files/nws-video1.wmv
147b105a666bb6a39ede3808f04ef78b	video files/nws-video3.wmv
a657bc8d88aa247524f4a8371e67e780	video files/nws-video4.wmv
2dc86420d0000e6811d3934362afd182	video files/nws-video6.wmv
2fe01d938a8f389568543f14233308f3	video files/plum.MOV
7162a639710142aad4f24c4012d3991e	video files/seedlings.MOV
507adcebbb57a813ecef189c0af0f7cd	video files/USGS-video1.mp4
524d60119de9a9d896f59fc1f449fd9f	video files/USGS-video3.mp4
f2468ce97fcd2af3184dd5703e8f24f7	video files/USGS-video4.mp4
214e87b5af32a0b6d1e791385f59527a	video files/USGS-video6.mp4

Appendix A4 – Email reference set

MD5	File
7ead9b1dea7f0d5a6d057d1dc79ef0e7	eml/Delete this mail.eml
b95768863d08e5f26f0f881aedb5f87c	eml/First email from yahoo to gmail (1).eml
161b1435b2ca3157e7e10b85f1395e7b	eml/First email from yahoo to gmail.eml
dd02553a83dd9829b4ae2f7bc3f08259	mbox/Brukerkonto1.email@gmail.com.mbox
de3a19cab697c2f7094a5b38182796d7	msg/ManyFields2.msg
d4dc90d5a2f8f5eca6ee5a879839e4d1	msg/ManyFields3.msg
bfbee2f23b7a725c87f9e3fdc333414e	msg/ManyFields4.msg
388e44b47eeb174e9c9b28390facc926	msg/ManyFields5.msg
f001c50dc4f6c798bcdc062d6d3fbf89	msg/Pink Note.msg
d584b2114c64dd140e784bdd43485360	msg/Regular.msg
954cbea6d8e99b8c414266926fd53c5a	pst/brukerkonto1_outlook.pst
fc0e789dc2408ff71587b01a5197db09	ost/Brukerkonto1.email@yahoo.com.ost

Appendix A5 – Archives reference set

MD5	File
f81e7aa7acb3fe9105a4c064030af377	Archives/archives with documents/7z_archive.7z
80c95b70032c2caf29e10db7745e198f	Archives/archives with documents/tar_archive.tar
17746dd75def22acab29c5d326b5bf6d	Archives/archives with documents/tar_gz_archive.tar.gz
14ec1134a41f34817a56a5491a55a831	Archives/archives with documents/wim_archive.wim
05a787a55383fe8b005ca39bd4703d88	Archives/archives with documents/zip_archive.zip
46018d02cbffef8f0e42304596672f54	Archives/Archives with images/Archive_7zip.7z
f0ca9907260522b9ba6c6c8656f036f9	Archives/Archives with images/Archive_tar.tar
39922700fe56378c5705c279ebb82e37	Archives/Archives with images/Archive_tar_bz2.tar.bz2
7204e658e493fb9636183d4f80dd5e0c	Archives/Archives with images/Archive_tar_gz.tar.gz
6eeea5e3f383fe9af46fd0040bcc91ed	Archives/Archives with images/Archive_wim.wim
f3bd1300411041234b3e20541bd10533	Archives/Archives with images/Archive_zip.zip
c724cd8088882cc25d63073ff1414210	Encrypted Archives/Document for encryption.zip
b8e6def9b5c469d81b9fbed9f9943e02	Encrypted Archives/Document for encyption 7zip.7z
9bf4698ed7503be4b4df373fe280d171	Encrypted Archives/Veracryptvolume
e1e7ece58fd27cc53e1926f691a38855	Encrypted Archives/BitlockerVHD.vhd
fa481ce2d61fef5cbf811a24b1e0f43d	Encrypted Archives/BitlockerVHD_Large.vhd
32a95417f8f9921265525c59034099e9	Encrypted Archives/LargeFileForEncryption.zip
6db90163ee34d5d3a0f487503020af90	Encrypted Archives/LargeFileForEncryption.7z
d71e94b14fb498e1181c7c36a5d62510	Encrypted Archives/Veracryptvolume_Large
95d1f19acd4a1ec7940873ebb0109ba7	cutty-sark.JPG
099ff446414d623109d719ed590f8412	doc_for_archive.docx
9ac20534486d249518f2aecce05ee34e	pdf_for_archive.pdf
07c5eda4d026a5dd59effc6ab09ba6dd	xlsx_for_archive.xlsx

Appendix A6 – Documents reference set

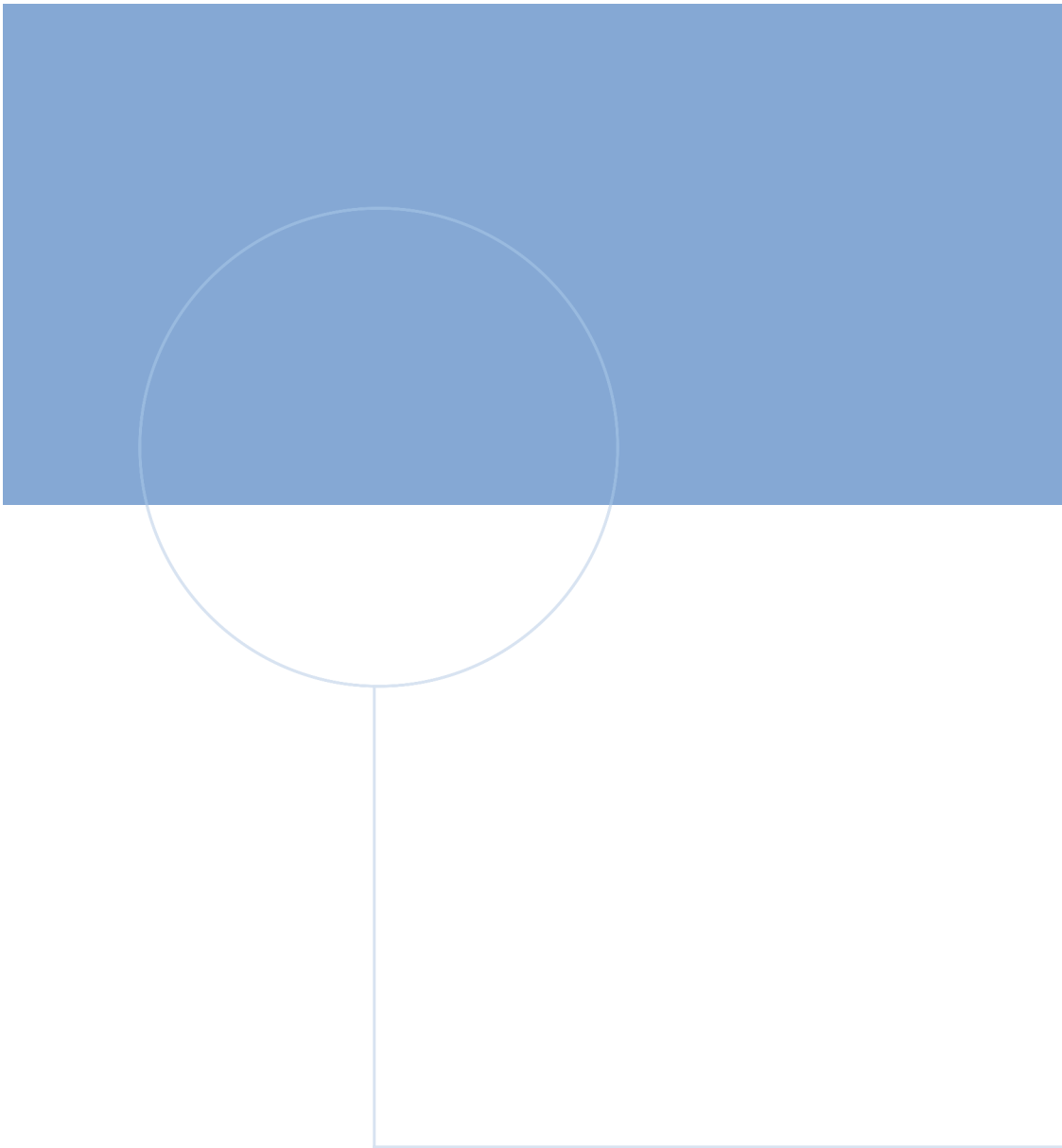
MD5	File
9f83b343869dc97e5bb1901effa8b9f8	xlsx_with_images.xlsx
e9ea6fb697d5048f244d288ca23eb02f	xlsx\gallopXLSX.tiff
d1fedbe48e875a328de535c9af8cbeba	xlsx\cutty-sarkXLSX.JPG
75a98db1351f9d0bf4f5b71c7a7a2e7f	xlsx\cactusXLSX.png
478837dd0ff326cb5e640912fea5ecf4	xlsx\bamboo-clump-anamatedXLSX.gif
3cddc6e794036580b954c98fd79c3468	xlsx\amalfiXLSX.bmp
243d15754ed5accdf72f620f53ecd204	rtf_with_images.rtf
4ca7ccdc77448646ba96d4d8b51eaff5	rtf\gallopRTF.tiff
77d7a8700572080ea475a1580a0dc8a2	rtf\cutty-sarkRTF.JPG
3032991ec2edec3c1a61498129aefd55	rtf\cactusRTF.png
6ab2205464aa30b3618afe5ab17a66c8	rtf\bamboo-clump-anamatedRTF.gif
c7553a3d9742eaddf36e4019555be7c8	rtf\amalfiRTF.bmp
9312e546c37925ac7ef837de9a42fe88	pptx_with_images.pptx
e8ed0df8ddeb718cd6cf0e2f42d3ef70	pptx\gallopPPT.tiff
88911a4747506fdb08bc68f2ce0b05c4	pptx\cutty-sarkPPT.JPG
5070978dcc7dd392856ce974358dfc8b	pptx\cactusPPT.png
caead00d61191bb782c6cc0d5d707434	pptx\bamboo-clump-anamatedPPT.gif
e4e79ff04fd43e18d73ebcce16a17b89	pptx\amalfiPPT.bmp
19fdbf2b04630d021948248d8eecabf9	PDF_with_images.pdf
e1f0c12c7936e78ad0a96cec91f419c6	PDF\gallopPDT.tiff
a3b9ff23e3306b37ceb3fac0bbeeaa29	PDF\cutty-sarkPDT.JPG
b17072ce6234b53a4957def7e2758daf	PDF\cactusPDT.png
08d1686819e502f62e1c0093fd537179	PDF\bamboo-clump-anamatedPDF.gif
89a03fc7f6b8a60f7fa10cde26c95f68	PDF\amalfiPDF.bmp
78a69d8993b4bfc94799b22f044d4631	odt_with_images.odt
7151353d153d951fa10f35f8b46e406b	odt\gallopODT.tiff
5338daaed6fef1cfecc77fefd5b21a71	odt\cutty-sarkODT.JPG
2a006eab11181665a73c3bd46382d8a3	odt\cactusODTODT.png
64527e92a25c99088efab1135d074481	odt\bamboo-clump-anamatedODT.gif
8256decb36e327ca50efaf6167a21ba3	odt\amalfiODT.bmp
a473c59ddf41660236288053d4cd6606	ods_with_images.ods
2f7b616b836f59b98337e7edb15c5631	ods\gallopODS.tiff
e6badf3b15d3d6c8185c0550648aa4ea	ods\cutty-sarkODS.JPG
9cf2c8ae9734c4b3a447f84873fa3dbb	ods\cactusODS.png
49d766865c95005daa35c7ddd43bd3ad	ods\bamboo-clump-anamatedODS.gif
35a8e240389ddf3a8f2f829bb9c738c4	ods\amalfiODS.bmp
712eea6d3a92af22f12c8128beec58d3	docx_with_images.docx
e003c5ec472e70be30178da6ca8afd52	docx\gallopDOCX.tiff
4b61bc2bfc99c269aae2feb33963c2d7	docx\cutty-sarkDOCX.JPG
e309c219b66210f6480e7bb1c537a57c	docx\cactusDOCX.png
1f031e15fda71b7754a727fd2c5ab223	docx\bamboo-clump-anamatedDOCX.gif
3c1d5096c9be5f70bf4c73eed014eec1	docx\amalfiDOCX.bmp

Appendix A7 – Documents reference set

Time	URL/activity	Browser	OS
24/03/2019 19:20:57	https://www.mozilla.org/en-US/privacy/firefox/	Firefox	Linux
24/03/2019 19:20:57	https://www.mozilla.org/privacy/firefox/	Firefox	Linux
24/03/2019 19:26:57	https://www.nist.gov/	Firefox	Linux
24/03/2019 19:27:33	https://www.swgde.org/	Firefox	Linux
24/03/2019 19:27:52	https://www.edrm.net/	Firefox	Linux
24/03/2019 19:28:06	https://www.google.com/?gws_rd=ssl	Firefox	Linux
24/03/2019 19:28:06	http://www.google.com/	Firefox	Linux
24/03/2019 19:28:06	http://google.com/	Firefox	Linux
24/03/2019 19:28:33	Digital - Google-søk	Firefox	Linux
24/03/2019 19:28:40	Forensics - Google-søk	Firefox	Linux
24/03/2019 19:28:47	Computer - Google-søk	Firefox	Linux
24/03/2019 19:28:53	Software - Google-søk	Firefox	Linux
24/03/2019 19:29:04	http://goo.gl/FrUcJM	Firefox	Linux
24/03/2019 19:29:05	https://unsplash.com/photos/p0j-mE6mGo4/download?force=true	Firefox	Linux
24/03/2019 19:29:05	https://goo.gl/FrUcJM	Firefox	Linux
24/03/2019 19:29:11	https://images.unsplash.com/photo-1550745165-9bc0b252726f?ixlib=rb-1.2.1&q=85&fm=jpg&crop=entropy&cs=srgb&dl=lorenzo-herrera-1383917-	Firefox	Linux
24/03/2019 20:30:25	https://chrome.google.com/webstore/detail/history-master/mkfgjjeggnmkbobjmelbjhdchcoadnin	Chrome	Linux
24/03/2019 20:30:38	https://chrome.google.com/webstore/detail/history-master/mkfgjjeggnmkbobjmelbjhdchcoadnin/related	Chrome	Linux
24/03/2019 20:31:05	https://www.nist.gov/	Chrome	Linux
24/03/2019 20:31:14	https://www.swgde.org/	Chrome	Linux
24/03/2019 20:31:27	https://www.edrm.net/	Chrome	Linux
24/03/2019 20:31:39	https://www.google.com/	Chrome	Linux
24/03/2019 20:31:49	Digital - Google-søk	Chrome	Linux
24/03/2019 20:31:56	Forensics - Google-søk	Chrome	Linux
24/03/2019 20:32:03	Computer - Google-søk	Chrome	Linux
24/03/2019 20:32:13	Software - Google-søk	Chrome	Linux
24/03/2019 20:49:06	chrome-extension://mkfgjjeggnmkbobjmelbjhdchcoadnin/option/index.html	Chrome	Linux
25/03/2019 20:37:28	History Master - Chrome Web Store	Chrome	OSX
25/03/2019 20:37:38	History Master - Chrome Web Store	Chrome	OSX
25/03/2019 20:38:08	National Institute of Standards and Technology | NIST	Chrome	OSX
25/03/2019 20:38:20	SWGDE	Chrome	OSX
25/03/2019 20:38:31	EDRM | Creating Practical Resources to Improve E-Discovery & Information Governan	Chrome	OSX
25/03/2019 20:38:43	Google	Chrome	OSX
25/03/2019 20:38:50	Digital - Google Search	Chrome	OSX
25/03/2019 20:38:56	Forensics - Google Search	Chrome	OSX
25/03/2019 20:39:02	Computer - Google Search	Chrome	OSX
25/03/2019 20:39:13	Software - Google Search	Chrome	OSX
25/03/2019 20:48:54	History Master	Chrome	OSX
25/03/2019 20:52:14	Firefox Privacy Notice — Mozilla	FireFox	OSX
25/03/2019 20:52:14	https://www.mozilla.org/privacy/firefox/	FireFox	OSX

Time	URL/activity	Browser	OS
25/03/2019 20:53:14	firefox history master - Google-søk	FireFox	OSX
25/03/2019 20:53:17	History Master – Get this Extension for Firefox (en-US)	FireFox	OSX
25/03/2019 20:53:46	https://www.nist.gov	FireFox	OSX
25/03/2019 20:54:02	https://www.swgde.org/	FireFox	OSX
25/03/2019 20:54:16	https://www.edrm.net/	FireFox	OSX
25/03/2019 20:54:41	Google	FireFox	OSX
25/03/2019 20:54:41	https://Google.com	FireFox	OSX
25/03/2019 20:54:47	Digital - Google-søk	FireFox	OSX
25/03/2019 20:54:55	Forensics - Google-søk	FireFox	OSX
25/03/2019 20:55:01	Computer - Google-søk	FireFox	OSX
25/03/2019 20:55:08	Software - Google-søk	FireFox	OSX
25/03/2019 20:55:19	https://unsplash.com/photos/p0j-mE6mGo4/download?force=true	FireFox	OSX
25/03/2019 20:55:19	https://goo.gl/FrUcJM	FireFox	OSX
25/03/2019 20:55:19	http://goo.gl/FrUcJM	FireFox	OSX
25/03/2019 20:55:29	lorenzo-herrera-1383917-unsplash(1).jpg	FireFox	OSX
25/03/2019 21:03:05	History Master	FireFox	OSX
25/03/2019 21:07:20	https://www.nist.gov	Safari	OSX
25/03/2019 21:07:50	https://www.swgde.org/	Safari	OSX
25/03/2019 21:08:15	https://www.edrm.net/	Safari	OSX
25/03/2019 21:09:15	https://Google.com	Safari	OSX
25/03/2019 21:09:35	Digital	Safari	OSX
25/03/2019 21:09:55	Forensics	Safari	OSX
25/03/2019 21:10:15	Computer	Safari	OSX
25/03/2019 21:10:30	Software	Safari	OSX
25/03/2019 21:10:55	goo.gl/FrUcJM (search)	Safari	OSX
25/03/2019 21:11:25	www.goo.gl/FrUcJM (search)	Safari	OSX
25/03/2019 21:12:15	goo.gl/FrUcJM (visit)	Safari	OSX
26/03/2019 19:15:33	history master chrome extension - Google-søk	Chrome	Windows
26/03/2019 19:15:39	History Master - Chrome Web Store	Chrome	Windows
26/03/2019 19:15:58	History Master - Chrome Web Store	Chrome	Windows
26/03/2019 19:17:46	https://www.nist.gov/	Chrome	Windows
26/03/2019 19:18:00	https://www.swgde.org/	Chrome	Windows
26/03/2019 19:18:14	https://www.edrm.net/	Chrome	Windows
26/03/2019 19:18:28	https://www.google.com/	Chrome	Windows
26/03/2019 19:18:37	Digital - Google-søk	Chrome	Windows
26/03/2019 19:18:44	Forensics - Google-søk	Chrome	Windows
26/03/2019 19:18:51	Computer - Google-søk	Chrome	Windows
26/03/2019 19:18:59	Software - Google-søk	Chrome	Windows
26/03/2019 19:22:24	History Master	Chrome	Windows
26/03/2019 18:57:04	https://www.mozilla.org/privacy/firefox/	Firefox	Windows
26/03/2019 18:57:05	Firefox Privacy Notice — Mozilla	Firefox	Windows
26/03/2019 18:59:07	https://www.nist.gov/	Firefox	Windows
26/03/2019 18:59:41	https://www.swgde.org/	Firefox	Windows
26/03/2019 19:00:05	https://www.edrm.net/	Firefox	Windows

Time	URL/activity	Browser	OS
26/03/2019 19:01:52	Google	Firefox	Windows
26/03/2019 19:01:52	https://www.google.com/	Firefox	Windows
26/03/2019 19:02:04	Digital - Google-søk	Firefox	Windows
26/03/2019 19:02:11	Forensics - Google-søk	Firefox	Windows
26/03/2019 19:02:18	Computer - Google-søk	Firefox	Windows
26/03/2019 19:02:25	Software - Google-søk	Firefox	Windows
26/03/2019 19:02:36	http://goo.gl/FrUcJM	Firefox	Windows
26/03/2019 19:02:37	https://unsplash.com/photos/p0j-mE6mGo4/download?force=true	Firefox	Windows
26/03/2019 19:02:37	https://goo.gl/FrUcJM	Firefox	Windows
26/03/2019 19:02:45	lorenzo-herrera-1383917-unsplash.jpg	Firefox	Windows
26/03/2019 19:12:49	History Master	Firefox	Windows
26/03/2019 19:33:15	https://www.nist.gov	Edge	Windows
26/03/2019 19:33:30	https://www.swgde.org/	Edge	Windows
26/03/2019 19:33:55	https://www.edrm.net/	Edge	Windows
26/03/2019 19:34:15	www.Google.com	Edge	Windows
26/03/2019 19:34:30	Digital - Google-søk	Edge	Windows
26/03/2019 19:34:50	Forensics - Google-søk	Edge	Windows
26/03/2019 19:35:10	Computer - Google-søk	Edge	Windows
26/03/2019 19:35:30	Software - Google-søk	Edge	Windows
26/03/2019 19:35:50	goo.gl/FrUcJM	Edge	Windows



NTNU
Norwegian University of
Science and Technology