Document Metadata

48436
Digital Forensics

Table Of Contents

1 Topic Outline

Digital Forensics is an important extension of forensic science. Regular forensics is, by definition, a discipline used to discover facts about what happened during a crime. It can be broken into many specialisations such as Pathology, Toxicology, Psychiatry, and more relevantly, Digital. Digital Forensics offers a way to discover information during a crime scene through recovery and investigation of digital based data found on devices such as Hard Drives, USB Drives, Mobile Phones, Smart Watches, and many, many more.

The investigation process itself may branch off into many different methods, however, this project will focus on Metadata investigation. Metadata itself is more of an umbrella term, with its simplest definition being "data about data". Structural and Descriptive metadata refer to information about the containers of data and the intellectual content of data respectively. Collectively, metadata can include authors, creation/modification date, file sizes, paths, and even the type of camera used on image files. Most computer files will have some sort of metadata, and gathering information through metadata is possibly one of the easiest ways to acquire initial bits of evidence.

Many metadata analysis/modification programs exist. Naming a few important ones include: OSForensics, HxD Hex Editor, and even the file command in Linux. A trivial use of metadata is evidence obfuscation. For example, a suspect may corrupt a file by changing its extension type in the file header using a Hex Editor such as HxD. By inspecting this file we could determine a mismatch between the extension in the file name and the "Magic Signature". Forensics experts can then revert the hex changes and make the file accessible, possibly uncovering some evidence. Malware is often masked as a different file type, inspecting its magic signature can help easily identify any possible malware. The list for the use of metadata goes on, and it is an extremely useful way of uncovering evidence or information.

This report will explore what metadata actually is, how it's used, how it can be extracted and how malware can be embedded into files. This report will contain an interview conducted with a Cybersecurity professional, addressing many questions in the digital forensics, and metadata fields. We will also explore sound business practices for securing metadata. Finally our reflections and conclusion will end the report.

The proposed report will delve into the definition, usefulness, and ways of metadata investigation. A thorough review of different types of malware, and their embedding techniques through metadata manipulation will be conducted. An in depth analysis of metadata types will be done, as well as demonstrations of metadata retrieval, modification, and analysis. Finally our proposed report will provide a recommendation into business practices to ensure the validity of all document metadata.

During the project, an interview will be conducted with a Cybersecurity professional, regarding metadata and its importance, all of which will be presented in the final report and presentation.

2 Schedule of Work

The schedule of work down below will be an outline into the tasks that will be needed to be done for the completion of the assignment. We intend to work on the tasks outlined below where we will be helping each other out in tasks so that we can all further our understanding on the topic of metadata. As such we have decided not to assign one individual to a certain task due to the collaborative effort.

Week 7:

- Contact CyberSecurity professional and set an interview date
- Research/literature review of metadata
- Gather real-world examples of metadata usage in investigations

Week 8:

- Further research into Windows based metadata analysis/modification tools
- Further research into Linux (Kali) based metadata analysis/modification tools
- Begin learning identified tools

Week 9:

- Perform and demonstrate metadata analysis with identified Windows and Linux tools
- Research different real-world malware and their embedding techniques (obfuscation etc)
- Attempt to demonstrate malware being embedded within a document
- Conduct interview with digital forensics professional

Week 10:

- Research into the common practices that security organisations use in relation to metadata
- Start preparing needed information for the presentation
- Consolidate all work into well-structured report (conforming to font size, spacing etc)
- Create and consolidate data into presentation

Week 11:

- Final edits made to report and presentation in preparation for submission.
- Final practice runs for presentation
- Submit final report

Week 12:

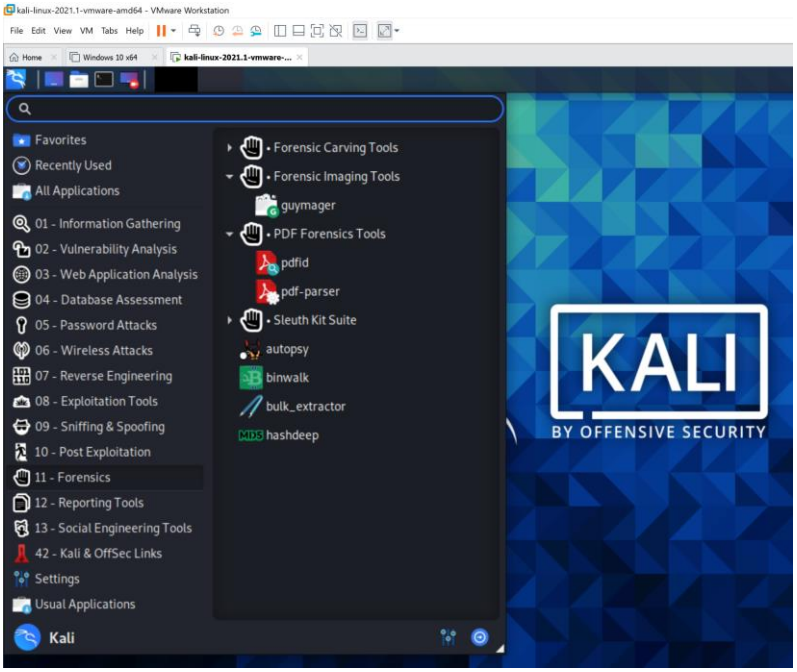
- Conduct presentation

3 Proof of Tools And Analysis

As explored throughout metadata is key in identifying critical pieces of information in the digital forensics process. Though gaining access and being able to use this metadata effectively is an important component and will determine the level of analysis we will be able to have on the metadata. In regards to some of the details of what will be covered through the final report submission we will look at using a whole range of tools (as per below. This will help with a further analysis of the metadata that will be presented in the final report and enable us to run and test some of the research we have done ourselves.

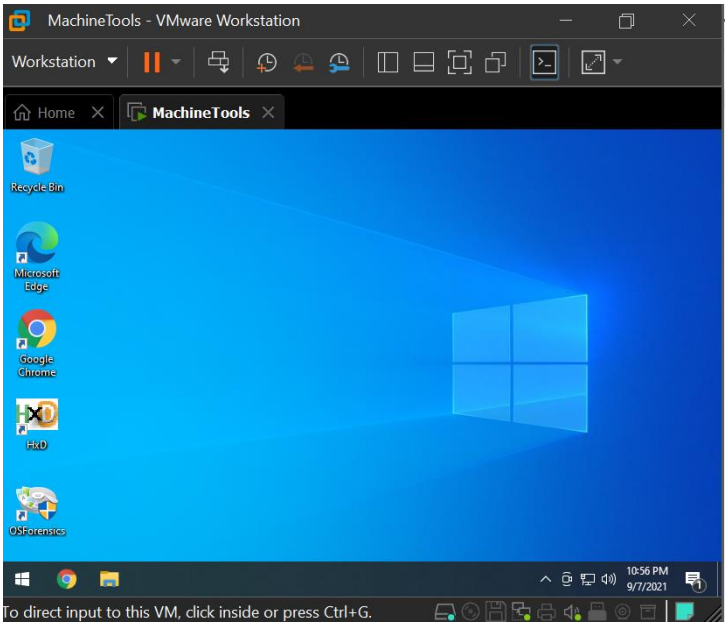
- Kali Linux OS

The use of Kali Linux is used often with digital forensics investigators as it offers a wide variety of tools that can be used, these are also specific to the digital forensics space, as seen below new versions of Kali offer specific forensics tools, and also have another section of Information gathering which is also useful for us in utilizing tools. These tools will be used throughout and we will look at each of the tools that are specific to metadata analysis and try to create some understanding of them (as planned in our week 7 work). Additionally we will be using Windows to the same capability, and research within its tools that can be used for metadata.



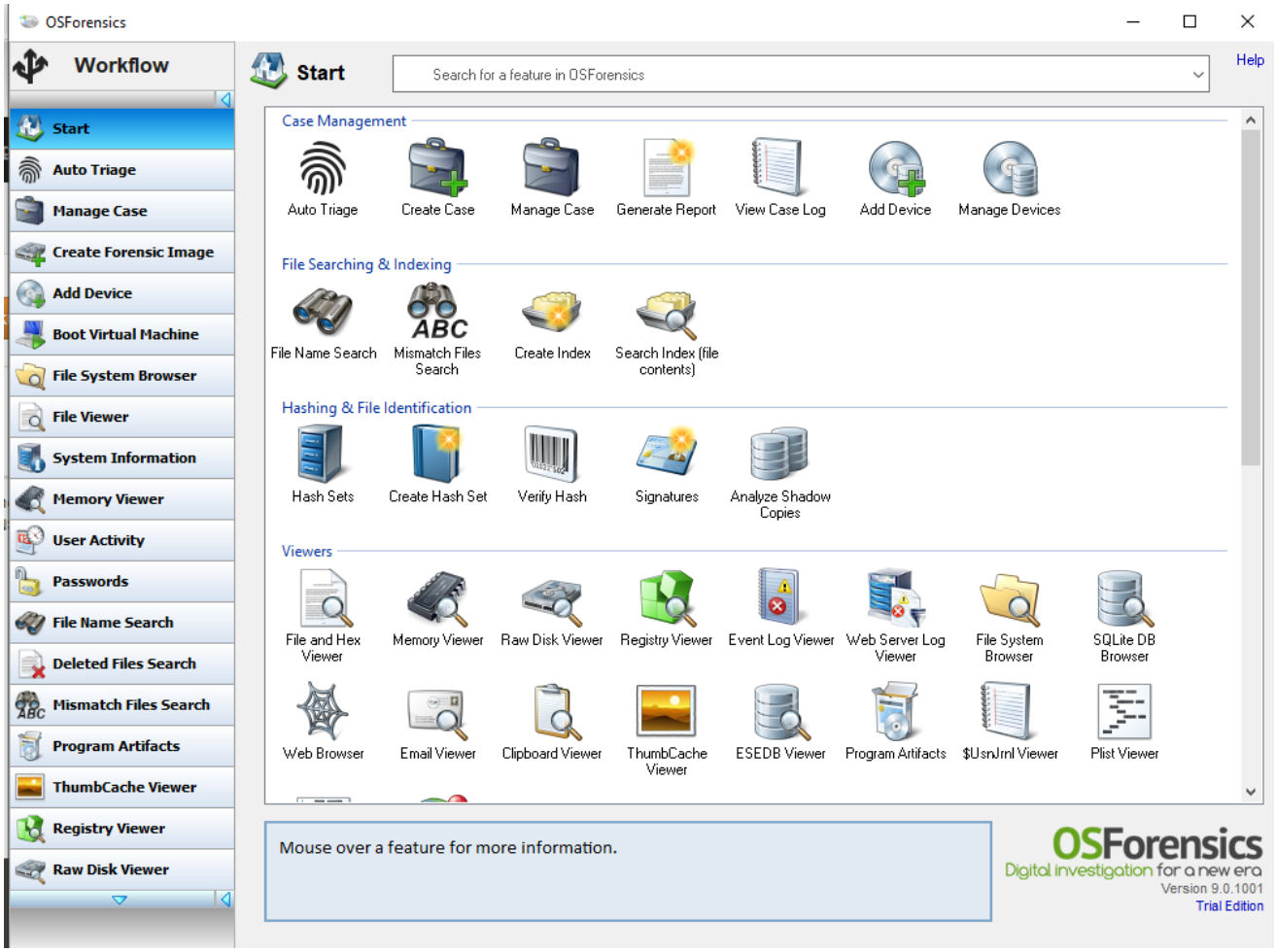
- Windows OS

We will also be using windows as another method of analysis metadata in our assignment for the purpose of seeing how some documents can execute scripts, but also using some of the basic tools windows has to analyse



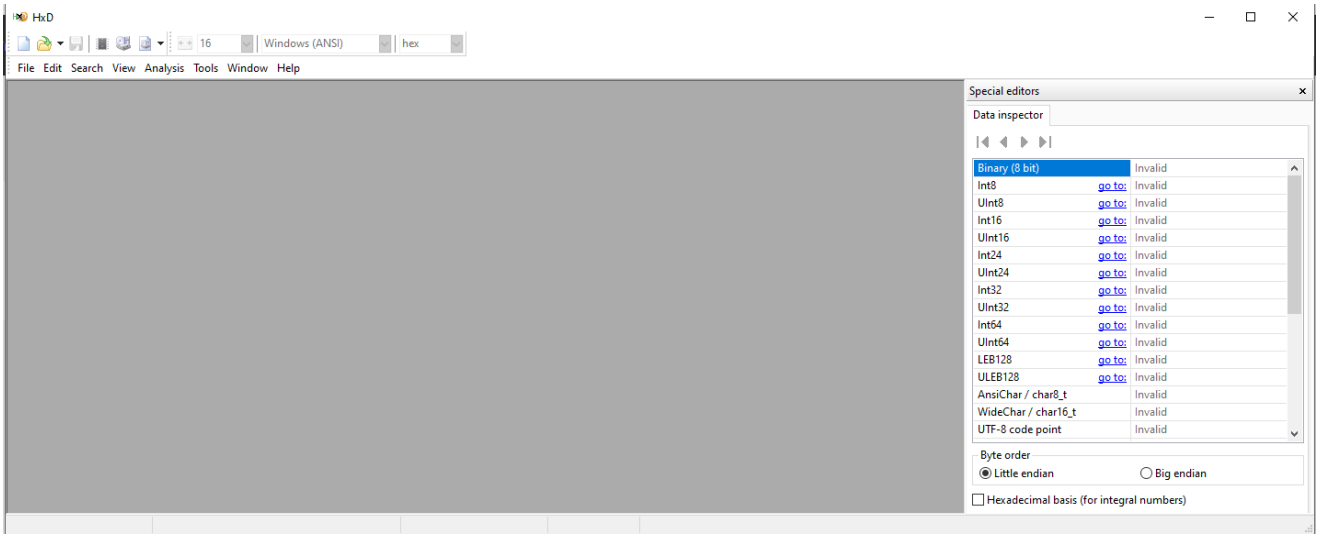
- OSForensics

OSForensics as seen in the labs is a tool that will be used for further understanding into documents and is useful in identifying metadata of files such as documents. This software lets you extract forensic evidence from computers quickly with advanced file searching and indexing and enables this data to be managed effectively. One of its useful features is also that it can potentially retrieve and recover deleted documents.



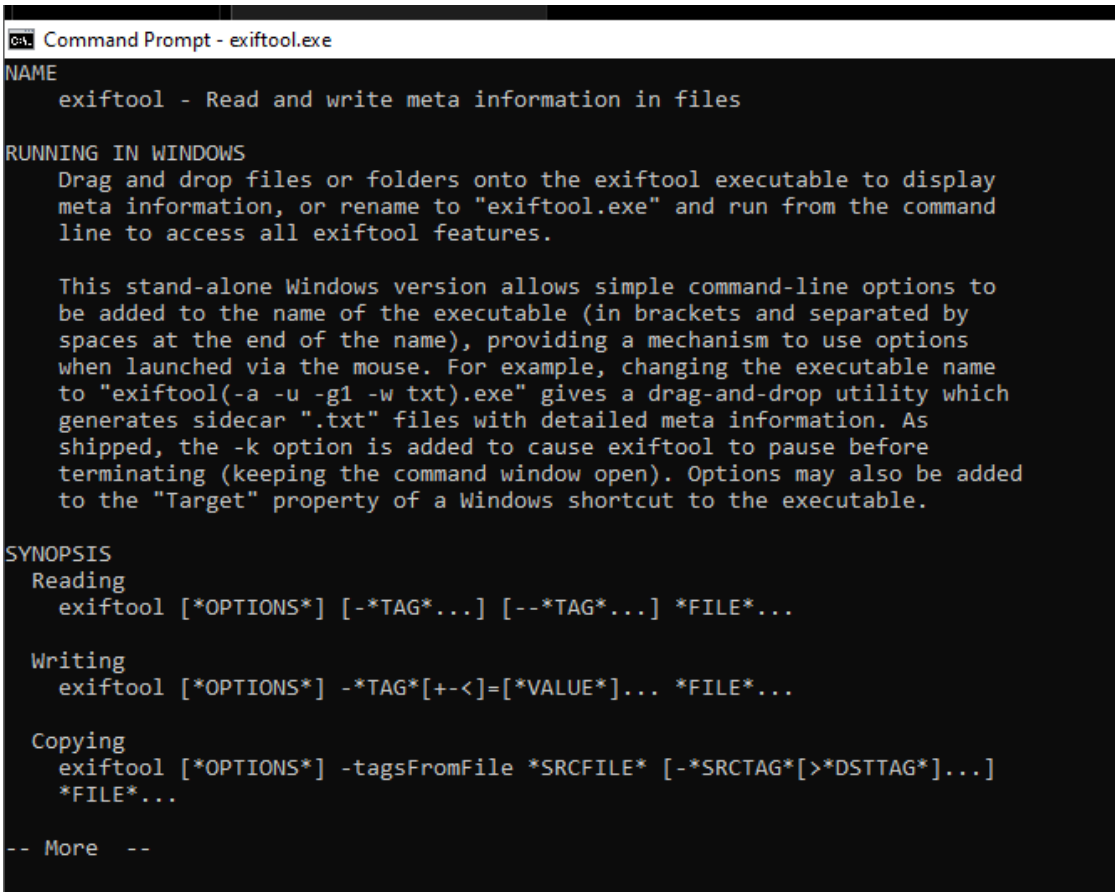
- HxD

The hex editor software opens any type of files and displays them byte by byte in hex values. Viewing the hex values are useful for analysing file signatures as viewing documents within our interest might have been altered. Additional advanced information can be found such as timestamps and identifying malwares embedded into files.



- ExifTool

ExifTool is a tool that is used for reading and writing metadata in various types of files such as documents, images, video and audio files. The output of running ExifTool will be the data of interest as returning values of dates, owner of the file and many more



- Oletools

This is a Python package utilising tools to extract metadata from OLE files. The Python tool is a great tool for malware analysis, forensics and debugging. It extracts and analyses Flash objects (SWF) that may be embedded in files such as MS Office documents (e.g. Word, Excel) and RTF, which is especially useful for malware analysis.

```
kali@kali:~/Desktop$ ole
olebrowse  oledir     olefile    oleid      olemap     olemeta    oleobj     oletimes   olevba
kali@kali:~/Desktop$ oleid Topic\ 1\ -\ Template\ and\ Instructions\ for\ Research\ Project\ \(1\).docx
oleid 0.60.dev1 - http://decalage.info/oletools
THIS IS WORK IN PROGRESS - Check updates regularly!
Please report any issue at https://github.com/decalage2/oletools/issues

Filename: Topic 1 - Template and Instructions for Research Project (1).docx
--------------------+--------------------+----------+--------------------------
Indicator           |Value               |Risk      |Description
--------------------+--------------------+----------+--------------------------
File format         |MS Word 2007+       |info      |
                    |Document (.docx)    |          |
--------------------+--------------------+----------+--------------------------
Container format    |OpenXML             |info      |Container type
--------------------+--------------------+----------+--------------------------
Encrypted           |False               |none      |The file is not encrypted
--------------------+--------------------+----------+--------------------------
VBA Macros          |No                  |none      |This file does not contain
                    |                    |          |VBA macros.
--------------------+--------------------+----------+--------------------------
XLM Macros          |No                  |none      |This file does not contain
                    |                    |          |Excel 4/XLM macros.
--------------------+--------------------+----------+--------------------------
External            |0                   |none      |External relationships
Relationships       |                    |          |such as remote templates,
                    |                    |          |remote OLE objects, etc
--------------------+--------------------+----------+--------------------------
```

- Pdfparser/pdfid.py

Python tools that are specifically analysing PDF document types. Pdfid allows you to identify PDF documents that contain for example JavaScript or execute an action when opened. PDfparser identifies the physical and logical structure of PDF Files which Load/parse objects and headers of PDF Documents. There is also a tool to create PDF malware documents.

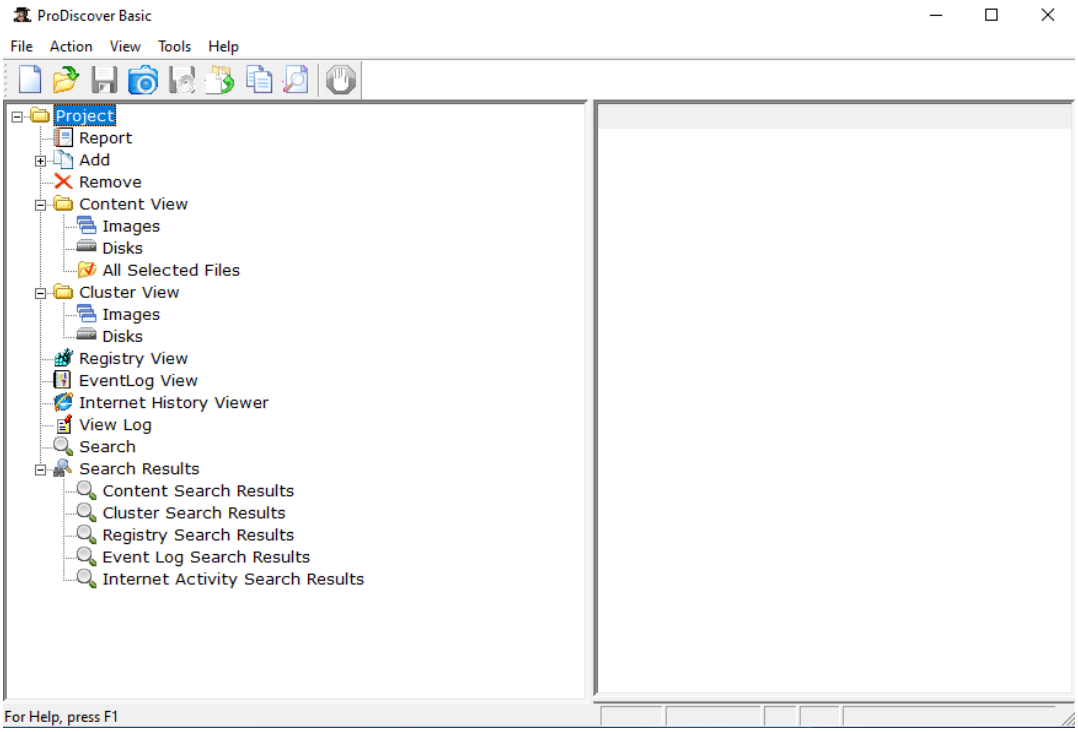
```
kali@kali:~/Downloads/pdfid$ python pdfid.py ../../Desktop/Lab\ 3\ AES\ Encryption\ and\ Decryption-1.pdf
PDFiD 0.2.8 ../../Desktop/Lab 3 AES Encryption and Decryption-1.pdf
 PDF Header: %PDF-1.6
 obj                    51
 endobj                 51
 stream                 42
 endstream              42
 xref                    0
 trailer                 0
 startxref               2
 /Page                   7
 /Encrypt                0
 /ObjStm                 8
 /JS                     0
 /JavaScript             0
 /AA                     0
 /OpenAction             0
 /AcroForm               0
 /JBIG2Decode            0
 /RichMedia              0
 /Launch                 0
 /EmbeddedFile           0
 /XFA                    0
 /URI                    0
 /Colors > 2^24          0
```

```
kali@kali:~/Downloads/pdfparser$ python3 pdf-parser.py
Usage: pdf-parser.py [options] pdf-file|zip-file|url
pdf-parser, use it to parse a PDF document

Options:
  --version             show program's version number and exit
  -h, --help            show this help message and exit
  -m, --man             Print manual
  -s SEARCH, --search=SEARCH
                        string to search in indirect objects (except streams)
  -f, --filter          pass stream object through filters (FlateDecode,
                        ASCIIHexDecode, ASCII85Decode, LZWDecode and
                        RunLengthDecode only)
  -o OBJECT, --object=OBJECT
                        id(s) of indirect object(s) to select, use comma (,)
                        to separate ids (version independent)
  -r REFERENCE, --reference=REFERENCE
                        id of indirect object being referenced (version
                        independent)
  -e ELEMENTS, --elements=ELEMENTS
                        type of elements to select (cxtsi)
  -w, --raw             raw output for data and filters
  -a, --stats           display stats for pdf document
  -t TYPE, --type=TYPE  type of indirect object to select
  -O, --objstm          parse stream of /ObjStm objects
  -v, --verbose         display malformed PDF elements
  -x EXTRACT, --extract=EXTRACT
                        filename to extract malformed content to
  -H, --hash            display hash of objects
  -n, --nocanonicalizedoutput
                        do not canonicalize the output
  -d DUMP, --dump=DUMP  filename to dump stream content to
  -D, --debug           display debug info
  -c, --content         display the content for objects without streams or
                        with streams without filters
  --searchstream=SEARCHSTREAM
                        string to search in streams
  --unfiltered          search in unfiltered streams
  --casesensitive       case sensitive search in streams
```

- ProDiscover

ProDiscover has capabilities to handle all aspects of an in-depth forensic investigation to collect, preserve, filter, and analyse evidence. The tool captures key evidence from computer systems. ProDiscover uses project-based reporting systems that log bookmarked data sets by the user



4 Interview Details

The interview with a Digital Forensics industry professional will take place on (Week 9), where we will work as a team to collaboratively make use of an online interview Question & Answer structure. Where we will be asking the professional questions which are specifically related to the digital forensics landscape, and we will be taking brief meeting notes during the session.

Down below is a list of questions we intend on asking during the session, we have intentionally made some of these questions vague to gain a broader understanding into the massive landscape of digital forensics. Additionally we will be working on adding additional follow up questions within the session dependent on answers as we feel this will allow for a more personalised interview which will allow for better explanation by the digital forensics .

Intended Interview Questions

1) What motivated you to work in the digital forensics field?
2) What is your favorite part of being a digital forensics professional?
3) Name an experience that you will never forget thanks to digital forensics ?
4) Could you name one time where you have been challenged working in the digital forensics field (specifically with an investigation) ?
5) What process do you follow as a digital forensics professional and would you recommend that process to new students within digital forensics?
6) What is the importance of metadata in forensics?
7) How important is it to confirm the validity of the metadata you collect from documents (and how do you check if they have been tampered)?
8) We are researching the topic of metadata as a professional, what tools do you use to analyse documents metadata ?
9) What are some other common tools you use at your workplace as a forensics professional ?
10) What advice would you recommend to future potential graduates looking at starting out in the digital forensics field ?