PUSL3119 Computing Individual Project

Project Initiation Document (PID)

Enhancement To Data Preservation with Foreda (Open-Source Tool for Data Preservation)

Supervisor: Mr. Chamara Dissanayake

Name: Jayasundara Bandara
Plymouth Index Number: 10747924
Degree Program: BSc (Hons) Computer Security Engineering

Table of Contents

CHAPTER 01 – INTRODUCTION

1.1 THE PROBLEM STATEMENT

Forensic science or criminalistic science is the association of criminal science and civil laws. In conclusion, forensic science includes all the factors of criminology such as analysis of DNA, blood pattern analysis, fingerprint analysis, questioned documents, toxicology reports, drug reports, pathology and various other different types of indirect and direct feasible evidence needed for a particular forensic investigation.

Forensic investigators collect, preserves, analyses the extracted evidence during the occurring of the investigation. After the processes of extractions, the forensic investigator faces certain problem. Such problems will be as invalidity of forensic discipline, insufficient validation methods, misleading testimonies etc. But out of these unidentified major topics one of the crucial problems in forensic science is data preservation. Investigators who interact with sensitive data faces a massive threat of evidence theft.

// For an example, testimony of forensic evidence can be misleading. In real life scenario few areas for potential errors exist, there have been incidents where the evidence was fabricated or the results that would have led to guilty convictions were concealed. Mistakes by the forensic investigator themselves can happen as well. Investigators can sometimes confuse or contaminate samples. //

Finally, when the examiners evaluate data collected from the devices the evidence presented to the law or a corporate can impact many lives and jobs. That is why the preservation processes in investigation should be up to date with the relevant security needed to hold the original data before and after processing. Preserving critical data during security incident is a must to gain a full consciousness about the overview of the incident and to establish the basement for further investigation about the threat. This data preservation doesn't mean only giving the necessary security measures but also successfully analyzing the incident by utilizing maximum data preservation standards to ensure that all the data relevant for the investigator is stored and captured safe in its original form rather than a broken format. Sometimes over role storing of rapid evidence in the database can corrupt the original format of the data. This may even lead to false accusations. The data entered simultaneously for the same database can jumble each other and make corrupted data.

Unlike physical evidence which can be collected and stored in a physical space the digital evidence exposes to a host of concerns. Volatility is one such concern, the digital evidence is volatile. This means it can change quickly and frequently. Any such change in the data can compromise the whole device and the data itself. Environment effects a device frequently, a change of the temperature can even fry out a digital device disks while tampering the data.

Human error is common in this scenario, in such cases where the failure to properly handle electronic devices that contain the evidence may also cause evidence to be damaged when analyzed. These are some other major problems in forensic data preservation.

1.2 EXISTING STATE OF THE PROBLEM

*"**How do they preserve?"***

Upon answering many questions related to evidence preservations talking about what exciting technologies are available to preserve is important. But the answer is complicated as it seems. Different investigators around the world uses different approaches in preserving these valuable data. Some may even use digital wallets. Research also suggests that some investigators store them physically in a physical space.
Also researches suggests that many investigators store digital evidence in electronic devices such as USBs. HARD DISKS etc.

1.3 PROBLEM SOLUTION

As for providing support for the forensic investigator in storing and preserving the data collected within an investigation, on or before analyzing the data/evidence a data preservation tool is meant to be build using python. The tool is called '*FOREDA',* a mixed word combined with two Spanish words as forensis and privada. Forensis is the Spanish word for forensics while as privada is the word in English for private or privacy.

Truthfulness privacy data market is a simple marketing tool used by businesses to encrypt and store sensitive data. It's a provident strategy used by organizations to stop data reduplications by increasing security measures in it. Even though many organizations are not used to these technologies. Yet, companies like Google and Facebook Inc currently uses this methodology to hide sensitive data. But also, these technologies are common in Block chains as the verifying method can potentially help people to identify the data, they obtain is exactly what they originally want.

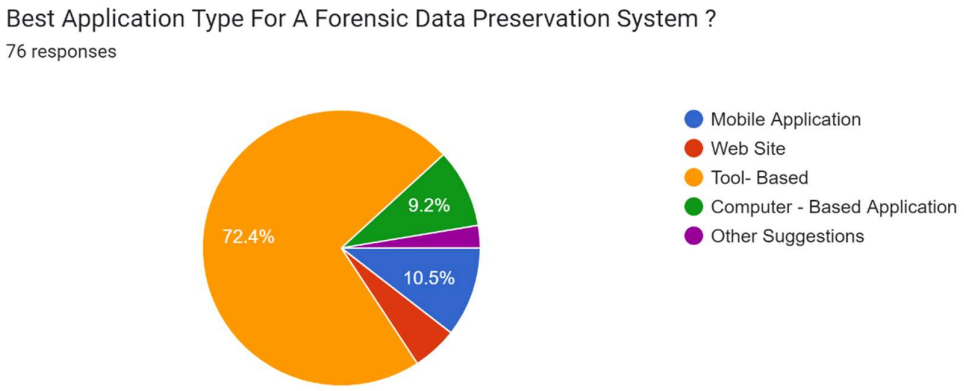
The tool '*FOREDA'* will introduce TPDM structure in its latex part. TPDM or truthfulness and privacy of data market is a structured homomorphic encryption and an identity-based signature format that simultaneously facilitates batch verifications, data processing and outcome verifications, while maintaining and continuing identity preservation and evidence confidentiality. This way the tool can encrypt huge batch files of evidence into an encrypted format. The tool will not only answer the data preserving question itself only. But also facilitates to help the investigator in certain evidence related matters as data reprocessing. *'FOREDA'* will carry a python written data reprocessing kit that will convert textual data to a comma-separated values file (.CSV).

This tool will also carry other sub kits like documentation clustering which will be essential for the investigator in data analysis process. This document clustering could cluster analysis textual documentations and extract topics, fast information retrieval and filter the document into a smaller format giving the user the requirements he needed only. Other sub filters will be as excel range filtering that will filter data in rows and columns and web scraping. Web scrapping is one of the exclusive kits that uses python bots to fetch analysis data on the internet compared to the relevant evidence.

Further hoping to add certain features like image forensic such as recovering corrupted images and enumeration directories like extracting the usernames and machine names of the evidence taken devices if time duration is applicable to be completed. And this is how the question for the major question of data preserving in forensics is answered. In the future development stage certain features will be added as well as reduced depending on the practicality.

1.4 BACKGROUND AND OUTCOMES

The proposed response for the question related to data preservation is a Python based tool. As forensics isn't a major subject studied by majority as well as forensic being rather a professional study that can impact many lives, the majority of public using this tool is rare. Only certain professionals may interact with this subject. Therefore, the necessity to build a website or a mobile application is not applicable for this subject.



1.4.1.0 Reddit survey on Application type to be built [Data taken from 10/10/2022 – 20/11/2022 by forensic studying community]

CHAPTER 02 – BUISNESS CASE

2.1 BUISNESS NEED

"The global forensic market size was valued at USD 4.30 billion in 2021 and is expected to expand at a compound annual growth rate (CAGR) of 10.2% from 2022 to 2030. The increase in the number of crimes globally is a major factor driving the industry's growth." [Referencing, A., 2020. Forensic Technology Market Size, Share & Trends Analysis Report. *Grand View Research,* 150(2020), p. 150.]

In forensic market, acquisition of evidence during an incident is a must to get a full incident overview and to establish a foundation for further investigation and threat eradication. One of the biggest needs for such introduced tool is originality. As described above one of the crucial factors that affect this market is data getting tampered and not existing in its original form. Which will lead to false accusations and misleading analysis. This may even be caused by anti- forensic tools. An article posted in Tandfonline will prove it,

[Referencing, Smith, A., 2007. Describing and Categorizing Disk-Avoiding Anti-Forensic Tools. *Tandfonline,* Volume IV, pp. 309-313.]

Annual reports suggests that the forensic science technologies are estimated to increase by 11% from 2021 to 2031. And out of that with the developing technological world the demand for data preservation will increase double the size of forensic industry. The moment an individual goes online they are in risk of getting exposed to someone else's desires. That's why an investigator being such an individual requires a tool to preserve.

2.2 BUSINESS OBJECTIVES

I. With this project the original data acquired will remain in its original format without getting tampered. This won't only save the investigators time in preserving but in a case if someone was falsely accused it will help these parties to prove themselves.

II. Rather than spending money on private cloud security software's or in fact in any other such wallet software's anyone can use this open-source tool to store the required data.

III. As the tool has truthfulness privacy data market system which is a tool that is used to encrypt sensitive data with its privacy and originality in financial businesses this tool can also be used to store data of other cyber security industries too.

IV. In forensics a 70% majority topic talked in this subject is preservation. And it is one of the costly topics talked. Using this tool institutions related to storing forensic data can reduce its costs and expenditures in preserving.

V. As the tool is meant to be built upon forensic digital evidence storing rules and regulations, the forensic investigators can ensure themselves in storing data safely.

VI. A main factor for the investigators in using this tool is time saving ability. The investigators will have the full required interface to store required data in separate related categorize as they wish.

CHAPTER 03 – PROJECT OBJECTIVES

3.1 DELIVERABLES

I. After clicking the tool, the user will be directed to a logging page where they can create a new account or log-in to an existing account. If the user goes through an existing account, they must perform the two- step verification. If the user forgets their password of an existing account, they will get a verification code for the email entered to reset the password.

II. Further the user should have their 06-pin code to unlock into their account. If the pin is incorrect the sessions will erase the entered username and password letting the user enter the relevant details again.

III. For security reasons as the tool is used by professionals to preserve sensitive data the username and password should carry upper case letters, characters, and numbers with a password length of more the 08 digits.

IV. The user will have a set of categorize with the storing category, sub kits as data reprocessing, documentation clustering, web scrapping, excel rang filtering etc.

V. Before using any sub tools mentioned above the user should agree for a privacy and security agreement as these tools are exclusive and can be used for any good or bad intentions.

VI. When data is entered to the storing category it will automatically encrypt them as for the truthfulness privacy data market system in its original form.

VII. Before accessing the storing category, the user should do the digital signature and author verification to identify if the user is the actual user accessing it.

VIII. The main tool will also carry a backup category in case the data in storing category gets exposed or deleted. This backup category won't be shown in the main page, but it will be hidden with encryption in the latex part of the tool. The user can either agree to backup or not in the agreement displayed before storing as they wish. And only the user can access it.

IX. After some time of not in use, the user will get automatically sign out as for security reasons.

X. The other tool kits in the main tool will be separated from data storing facility with separate encryptions in a user efficient manner.

3.2 DELIVERABLE'S FEASIBILITY

I. Having different accounts to sign in will be a main strength of the tool as many accounts can be made by one user as well as it prevents intruders from logging-in to someone else's accounts.

II. The two- step verifications will identify if the user is the actual user, they say is.

III. Unlike in social media applications and tools having a mixture of characters and numbers in the username can be helpful as it is a tool used by professionals for their day-to-day work.

IV. The sub tools mentioned in the introduction are rare yet dangerous tools. Without proper verifications and agreements any user can use these tools for ill intentions.

V. Having separate folders for data may prevent data from getting corrupted.

VI. As this tool carry's sensitive data having backups will be helpful in case of an emergency. But to backup or not to will be a user's choice. And hiding and separately encrypting these backups will be a strength of the tool.

VII. Having sessions to timeout can be important as if the owner leaves the computer opened with the tool anyone could access the data and other tools.

CHAPTER 04 – LITERATURE REVIEW

4.1 DOMIAN OVERVIEW

The project domain is selected on forensic basis subject while answering to a main question arose during the forensic before and after analysis, which is data preservation. Data preservation can be a crucial task and the idea to build this project as a computer-based tool is that the need for a data preservation tool is only required for a particular party. And rather than a mobile phone application technology which get exposed to malicious intentions more often that a computer, computer tool based could give more security for such sensitive data.

Securing evidence that the investigator retrieved is an important factor in an investigation. If wrong people retrieved such digital data, they could cause a major danger not only for the investigator but also to many other people. There have been such reports of data getting theft by the investigators by third parties. Reports also says that these third parties used the data to blackmail individuals for money. A new article posted in *the mirage* mentions about a Brisbane man getting arrested using stolen data to defraud hundreds of victims for tens of thousands of dollars.

[Referencing, The Mirage, 2022. Brisbane Man Arrested For Using Stolen Data To Commit Widespread Fraud, Brisbane: The Mirage.]

Meanwhile explaining how much precautious the evidence collected in a forensic crime scene, it is also important to know how important the evidence is for an investigation. The official forensic justice page of the United States government explains the value of the forensic data by stating that forensic science is a critical element in serving criminal justice., the forensic scientist will examine and analyze such evidence from a crime scene to give common objectives to assist the investigation by serving justice to the crime while absolve an innocent person from defraud.

[Referencing, United States Government, 2022. *Justice Forensic Science.* [Online]
Available at: https://www.justice.gov/olp/forensic-science
[Accessed 10 November 2022].]

4.2 EXISTING SYSTEMS

Moreover, many forensic scientists have come up with many solutions to preserve data. Some may even use cloud software's because of the ability to access data at any time anywhere. But the fact these users doesn't understand is that the evidence collected is not a normal form of personal data they can store in a cloud software or a wallet, these are precautious data. Throughout years some others came up with other software solutions from IOT forensics, Mobile Forensics, Web- based forensics etc. In this sub-chapter such developments will be clearly talked, appreciated, and criticized in general.

Cloud security is an intelligent solution, in recent years two engineers 'Mehran Pourvahab, Gholamhossein Ekbatanifard' proposed a digital data forensic preservation system in Iaas could environment using elements as SDN and blockchain technologies. In this system the evidence is collected and stored in the blockchain distributed among many peers. To protect the system from unauthorized users its is protected by Secure Ring verifications. And it contains Harmony Search Optimization algorithms to protect the cloud with many other features. The state of project is unknown yet, but it is one of the most appreciated technologies and the foundation to build *FOREDA*.

[Referencing, Pourvahab, M. & Ekbantanifard, G., 2019. Digital Forensic Architecture For Evidence Collection and Provenance Preservation in Iaas Cloud Environment. *IEEE Xplore,* Volume 7, p. 7.]

However, there are some downsides for this solution as internet connection is required to store data and cloud solutions likely have 20% chance of getting hacked than a computer-based tool says blogs.

[Referencing, Morgan, N., 2021. *Triskelelabs.* [Online]
Available at: https://www.google.com/amp/s/www.triskelelabs.com/blog/cloud-cyber-attacks-the-latest-cloud-computing-security-issues%3fhs_amp=true
[Accessed 20 November 2022].]

IOT forensics is another such co- related subject for data preservation. Internet of things getting cyber attacked had gradually increased recently. A research paper about a framework for internet of things- based block chain system has been introduced in Wiley online library. The framework here aims to integrate blockchain system in IOT to overcome problems faced by investigators by enhancing the ability to preserve. In conclusion block chain is the structure that stores data or blocks in several databases knows as the chain in a network connected through peer to peer.

[Referencing, Randa Kamal, E. E.-D. H. N. E.-F., 2022. Forensic Chian For Evidence Preservation System. *Wiley Online Library,* 34(21), p. 34.]

4.3 TECHINICAL ANALYSIS

- A published article about python mentions that python is the fastest growing language. In fact, python is simple, dense, updating and complex all in one package. Python also helps other high- level languages, network and internet programming, integrity of components, database programming, GUI, System programming etc. However, even though website backend development using python has disappeared, with the growing demand for big data the demand for python increased as well. And for open-source developments python is the best choice.

[Referencing, Srinath, K. R., 2017. Python - The Fastest Grwoing Programming Language. *ACADEMIA,* 04(12), p. 357.]

- The selected software technology for the development is PyCharm. It is one of the most famous python coding communities.

[Referencing Islam, Q. N., 2015. Mastering PyCharm. *Google Books,* Volume 10, p. 100.]

4.4 REFLECTION

Hence after a successful literature review investigation, studying existing problems about forensics in articles and how evidence can impact lives, a clear understanding about the risk of developing such proposed system is understood. While further studying the lesser number of system solutions for forensic data preservations a clear understanding about how to improve the proposed tool is taken. And these existing tools for preservation is well admired for the insights they gave in developing *FOREDA.* Hence after studying the technologies and literature review, hoping to make this project a success.

CHAPTER 05 – METHOD OF APPROACH

5.1 WORK THAT HAS BEEN DONE

- Identified the problems in forensic.
- Read articles related to data preservation in forensics.
- Identified what and how TPDM system works and where it is used.
- Studied python in depth.
- Talked to forensic scientists from social media platforms about the project and how to improve it.
- Understood some strengths and weaknesses in the existing projects.
- Documenting the requirements and features of the system planned to be built.
- Setting up the required physical and non-physical environments needed to make this tool a success.
- Setting the backing up parameters needed in an emergency. (e.g., GitHub)

5.1 WORK TO DO

- Development stage is estimated to start in December first week.
- During development stage, further research on existing preservation systems will be done while identifying ways to make FOREDA better.
- Getting professional help in studying more about python directories.
- Making of other related documentation reports to show the current states of the project.
- After development stage, testing the system in and out.
- Identifying any missed places related to the project and taking necessary steps to achieve the common goals.
- Other.

CHAPTER 06 – INITIAL PROJECT PLAN

6.1 DEVELOPMENT PHASE

- Planning the process – Clearly identifying and writing down the need for a tool with its purpose. While writing down the specifications and security details needed to achieve these common purposes with future implementations.

- Setting up environment – Identifying the relevant coding scenarios while installing and updating the relevant software's needed to build the tool '*FOREDA'*.

- Creating the tool (First stage) – With identified resources and studied research the databases related to logging's will be designed as per the requirements. Later the front-end for those user loggings will be designed as for the back end.

- Creating the tool (Second stage) – After the first level designing the second stage will carry the TPDM system. It will be built using python separately and inbuilt into the main tool after tests.

- Creating the tool (Third stage) – After the second stage this stage will carry on building all the separate encryptions needed for the tool's features.

- Creating the sub-kits – In this stage all the other related sub-kits needed for the investigator for data analysis will be created and added to the tool separately. Depending on time new features may or may not be added.

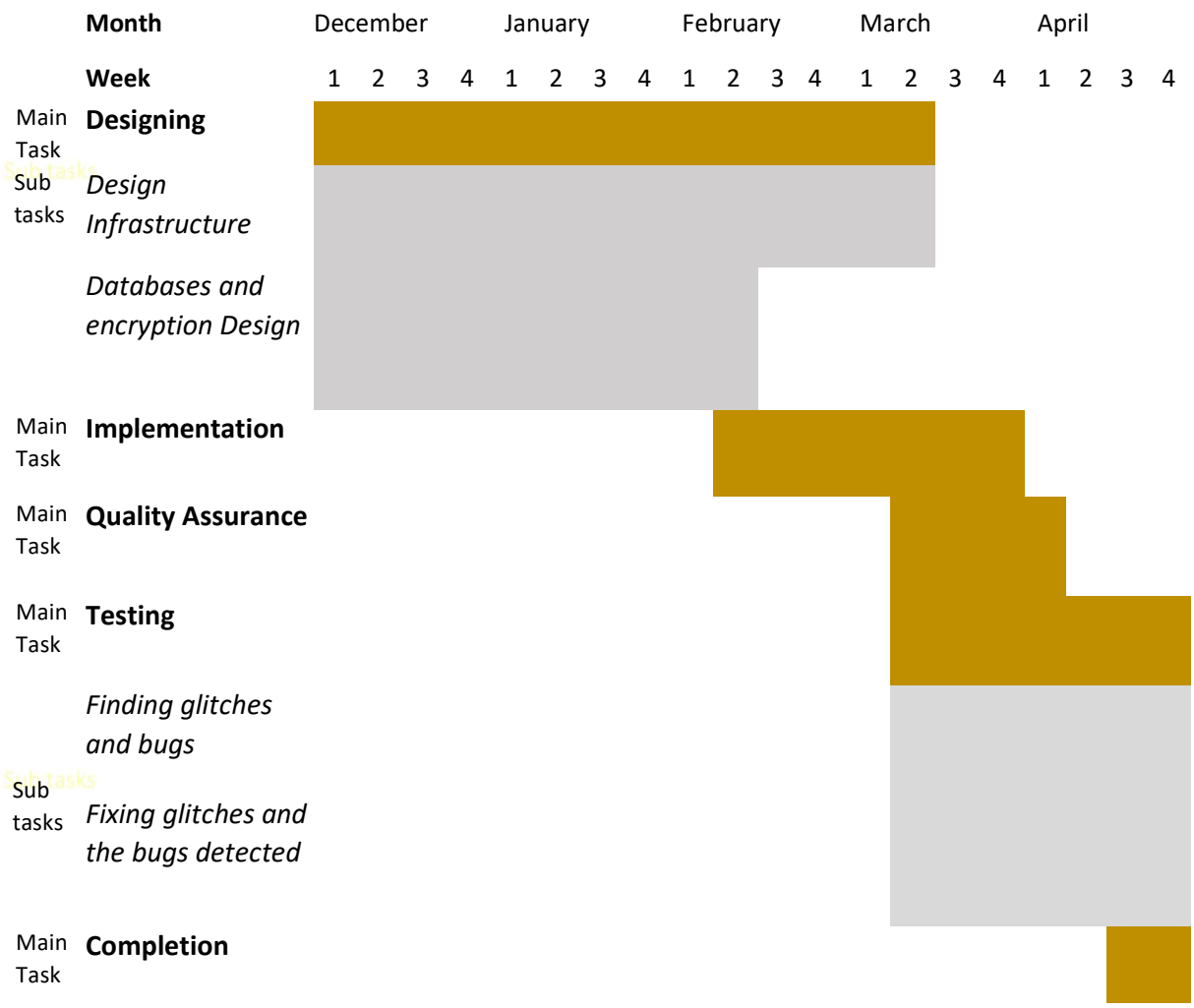
- Backing-up stage – Before the final connection of the creating tool stage all the python coded separate designs will be backed up for security assurance.

- Connecting stage – This stage is a long stage as per all the related python coded kits and encryptions will be added into the main tool.

- Outcome and testing stage – The final tool will be tested and penetrated with different tests to find out its weaknesses and strengths.

- Error free zone- During this stage all the errors and bugs will be fixed while exposing this tool for the industry.

6.2 TIME DURATION



6.2.1.0 Time duration Gantt chart

CHAPTER 07 – RISK ANALYSIS

7.1 POTENTIAL RISK

- One of the main risks that could arise during development stage is inability to complete the project before the due date.
- Inability to do tests before presenting the final project will lead to inaccurate data analysis.
- Less programming resources to find unknown factors related to coding during development stage.
- Technological failures related to computers used to build the tool may lead to data getting erased.
- Technological issues with software's used to build the tool can occur too.
- Getting frauded by online scams when asking for any social media help related to the project.
- Misleading information from professionals, articles and research papers may lead to an unpractical system.
- Getting work done stolen by other parties.
- Others.

7.2 RISK MANGEMENT MECHNISM

- Saving and backing up work of each day during the developing stage.
- Having a proper schedule to do related work for developing before exceeding the due date.
- Testing every new build stage regularly to find any bugs and fixing them.

- Testing the final tool again and again to ensure it's working successfully as for the requirement.
- Checking the physical and non – physical equipment's needed for the development before using them. (e.g., Computers, Software's etc.)
- Correctly justifying the information taken from professionals, articles, and research papers before using them.
- Understanding the objectives clearly in each stage of the development to stop the project from changings its purpose.

CHAPTER 08 – ARCHITETURAL DIAGRM

8.0.1.0 Simple Architectural Diagram

REFERENCES

References

A., 2020. Forensic Technology Market Size, Share & Trends Analysis Report. *Grand View Research,* 150(2020), p. 150.

Smith, A., 2007. Describing and Categorizing Disk-Avoiding Anti-Forensic Tools. *Tandfonline,* Volume IV, pp. 309-313.

The Mirage, 2022. Brisbane Man Arrested For Using Stolen Data To Commit Widespread Fraud, Brisbane: The Mirage.

Srinath, K. R., 2017. Python - The Fastest Grwoing Programming Language. *ACADEMIA,* 04(12), p. 357.

, United States Government, 2022. *Justice Forensic Science.* [Online]
Available at: https://www.justice.gov/olp/forensic-science
[Accessed 10 November 2022].

Pourvahab, M. & Ekbantanifard, G., 2019. Digital Forensic Architecture For Evidence Collection and Provenance Preservation in Iaas Cloud Environment. *IEEE Xplore,* Volume 7, p. 7.

Randa Kamal, E. E.-D. H. N. E.-F., 2022. Forensic Chian For Evidence Preservation System. *Wiley Online Library,* 34(21), p. 34.

Islam, Q. N., 2015. Mastering PyCharm. *Google Books,* Volume 10, p. 100.

Morgan, N., 2021. *Triskelelabs.* [Online]
Available at: https://www.google.com/amp/s/www.triskelelabs.com/blog/cloud-cyber-attacks-the-latest-cloud-computing-security-issues%3fhs_amp=true
[Accessed 20 November 2022].

GeeksforGeeks, 2021. RSA Digital Signature Scheme Using Python. *GeeksforGeeks,* p. 20.
Hendricks, Beth, 2020. *Issues in Digital Evidence: Risks, Prevention and Protection.* [Online]
Available at: https://study.com/academy/lesson/issues-in-digital-evidence-risks-prevention-protection.html
[Accessed 20 09 2022].

Humbolt State University, 2022. *Creating Custom Tools with Pythom Scripts.* [Online] Available at: http://gis.humboldt.edu/olm/Courses/GSP_318/07_2_ArcGIS_CustomTools.html

Palter, J., 2021. Securing, Keys, Assets & People - Blog. *Preserving Digital Evidence the Right way: Your 10- Step Guide,* p. 6.

Policies, S. o. U. g., 2022. Forensic Evidence Policy. *Forensic Evidence Policy,* 10(12), p. 10.

Simon, M., 2020. 3 Methods to Preserve Digital Evidence for computer Forensics. *3 Methods to Preserve Digital Evidence for computer Forensics,* p. 10.

Tim Grossmann, 2022. *How to Build a Bot and Automate your Everyday Work.* [Online] Available at: https://www.google.com/amp/s/www.freecodecamp.org/news/building-bots/amp/

Unknowm, 2021. *Turtorialspoint: Pyhton Forensics Tutorial.* [Online] Available at: https://www.tutorialspoint.com/python_forensics/index.htm [Accessed 15 October 2022].

Wei Wang, 2021. *Hindawi.* [Online] Available at: https://www.hindawi.com/journals/complexity/2021/5536326/ [Accessed 23 September 2023].

Wilson, J., 2017. *SmartData collection: What are the Limits of Forensic data Retention?.* [Online] Available at: https://www.google.com/amp/s/www.smartdatacollective.com/what-are-the-limits-of-forensic-data-retention/amp/ [Accessed 08 October 2022].