A Digital Forensic Investigation of the Presence of Personally Identifiable Information (PII) in Refurbished Hard Drives

Robinson Tombari Sibe[1,*], Blossom U. Idigbo[2]
[1]Rivers State University, Nigeria / Digital Footprints Ltd, Nigeria
[2]Digital Footprints Ltd, Nigeria
Emails: robinson.sibe@ust.edu.ng; blossom.idigbo@digitalfootprints.ng

Abstract

The last decade has seen a massive explosion of data, with a lot of Personally Identifiable Information (PII) flooding devices and the cyberspace. This has necessitated the growing call and global awareness for data protection, to ensure the responsible use of data, protect the privacy of data subjects, and prevent crimes such as identity theft and cybercrime. This paper investigated the presence of residual data and Personally Identifiable Information (PII) in refurbished hard drives bought from a retail shop. The study leveraged digital forensic tools to perform data recovery on refurbished hard drives, and analyses for presence of PII. The study adopted a modified form of the steps in Digital Investigation outlined by NIST IR 8354. Result of this study showed that one out of the 3 hard drives that were reportedly formatted and sanitized by the vendors had residual data with PII. Data recovered includes 28691 files with size on disk as 152.20GB, including PII and sensitive data. Digital Forensic tools used for this study includes EaseUS Data Recovery Wizard and Autopsy. The findings of this study are quite relevant to current studies in privacy and data protection, including recent legislations such as Nigeria Data Protection Act (NDPA), General Data Protection Regulation (GDPR), and others. The paper also presents a comprehensive and forensically sound software-based methodology focused on the recovery of deleted data from hard drives.

Keywords: Data Recovery; Privacy; Data Protection; NDPA; NDPR; GDPR; PII; Digital Forensics; Information Governance; Data Governance

1. Introduction

The last few decades have seen a massive explosion of data, including Personally Identifiable Information (PII). This has necessitated growing calls for responsible use of data and data protection. Across the globe, different jurisdictions have come up with different legislations. For instance, the EU released the General Data Protection Regulation (GDPR) in 2018. Nigeria released the Nigeria Data Protection Regulation (NDPR) in 2019, and only recently the Nigeria Data Protection Act in 2023 [23]. In Egypt, the Data Protection Law (law no. 151 of 2020) was passed [38]. These legislations all put greater responsibility on data processors and data administrators for data protection [39]. This study was designed to investigate the presence of PII and sensitive information on hard drives bought from a retail computer shop. The research leveraged data recovery tools to recover deleted data from the hard drives.

In an era dominated by digital information, the significance of data recovery cannot be overemphasized. Hard drives, ubiquitous in both personal and professional realms, serve as the repositories of our digital lives. The researchers embarked on a comprehensive journey into the intricate realm of data recovery in a forensically sound manner. The high cost of brand-new hard drives has meant many now resort to sales and purchase of refurbished hard drives. As the reliance on refurbished drives continues to soar, the accidental loss of crucial files poses a constant threat. Acknowledging this reality, this research unraveled the complexities of data recovery methodologies, offering insights, best practices, and solutions to navigate the challenges of recovering lost, deleted, or corrupted data. The study explored the intricacies of file systems and dissected the functionalities of specialized recovery tools.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

Data recovery is the process of restoring data that has been lost, deleted, corrupted or damaged from a storage medium. It can be performed on different types of storage media, such as hard disks, flash drives, memory cards, optical discs, and cloud services. Data recovery is a complex and specialized field that requires advanced tools, techniques, and expertise to achieve successful results depending on the cause and extent of data loss. Some common scenarios that require data recovery are accidental deletion, formatting, malware infection, physical damage, logical errors, and encryption [45].

Refurbished hard drives are previously used storage devices that have been reportedly worked upon and sanitized and can be purchased from retail stores or online. Following the rising cost of brand-new hard drives, most users are turning to refurbished hard drives. However, this is not without challenges. Apart from the challenge of durability - since they are not bought brand new - a major challenge with this is the fact that while it is common for the vendors to claim that the hard drives were sanitized, there are concerns that sensitive personal information, previously deleted, could be recovered from these hard drives. Using expert tools and skillset, deleted data can be recovered, thus leading to a potential breach of privacy, and criminal schemes such as identity theft [31].

Identity theft can have a wide range of consequences. These consequences can include gaining employment, obtaining credit card information and bank accounts, creating new ones, collecting loans, and exploiting the victim's name to buy prescription drugs and health insurance. Experts have noted that every year there are more than 50,000 individual personal data breaches in the US alone and this is because 87% of people leave personal information exposed. Research conducted by IBM showed that inappropriate disposal plays a role in 58% of data breaches [27]. This covers both electronic and physical disposals. For example, a user might dispose a computer that contains sensitive data or only erase files without permanently deleting them. This study investigated this, to show the extent of PII found in hard drives that have been reportedly refurbished and formatted for resale.

2. Literature Review

A. Data Storage and Lifespan

Recent research suggests that SSDs are primarily replacing secondary storage. However, HDDs still offer longer lifespan and are more reliable for storing data over an extended period [26]. As hard drives become more expensive, particularly, in economies with challenging macroeconomic indices, many users (individuals and SMEs) are turning to refurbished hard drives – either by selling used hard drives or buying refurbished hard drives. Refurbished hard drives (HDDs) are characterized by their affordability, large storage capacity, and relatively reliable data access within a 1-3year timeframe. Their ubiquity is further amplified by the growing emphasis on electronic device recycling, often leading to used computers and their HDDs entering secondary markets. However, in certain usage environments, HDD lifespan can exceed 1-3 years, increasing the likelihood of multiple ownership changes and raising concerns about data security and potential residual information leaks.

The reliability of hard disk drives (HDDs) is dependent on the drive construction, as well as the operational and environmental conditions in which the drive is used. Self-monitoring, analysis, and reporting technology (SMART) continuously provides attribute information on HDD usage and degradation characteristics [34]. Advanced data analysis techniques using artificial intelligence and machine learning algorithms have been developed to predict hard drive failures and classify remaining lifetimes based on S.M.A.R.T. features, with Random Forest Classifier achieving up to 94% accuracy [48].

Despite the environmental merits of device reuse, particularly for hard drives, concerns regarding data privacy remain paramount. Investigations have revealed an alarming prevalence of secondhand computers harboring sensitive information such as social security numbers, credit card data, trade secrets, medical records, and financial transactions. For instance, a researcher documented an abandoned HDD from an ATM, still containing complete transaction histories [24]. The data retrieval employed readily available tools, highlighting the vulnerability of such devices.

A pervasive misconception among users is that formatting a drive using standard Operating System tools (e.g., Microsoft Windows) effectively erases data. Contrary to this belief, secure data deletion necessitates overwriting each bit with a "0," or a “1,” mimicking the write process in duration. Alternatively, flagging the space as free allows new data to overwrite it eventually. The latter method, preferred for its speed and convenience by users solely seeking additional storage space, leaves data vulnerable to recovery. There are several software programs capable of data recovery through methods such as scanning unallocated space (slack space) [13]. Techniques that are more sophisticated involve platter removal and specialized hardware scans for data remnants.

B. Data Recovery Methods

Data recovery methods encompass both physical and logical approaches to salvage inaccessible data from storage media. These techniques range from simple backup and restoration to advanced forensic methods, addressing various causes of data loss [14]. Data recovery is an important skill set in digital forensics. It is a common practice

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

that cybercriminals may deliberately delete evidence as an anti-forensic practice. It is also possible that data may have been accidentally deleted, or the disk drive itself may have been damaged. Therefore, the need to recover deleted or damaged data has driven researchers to develop tools and methods for the data recovery process.

When a file or data is deleted using regular methods (for example through logical deletes in the Windows Operating System), such files may still be recovered even after the recycle bins have been emptied. Deleted files could still be found in slack space or unallocated space, and this could be extracted in a forensically sound manner. The factors affecting the success of the data recovery process may include the data type, the condition of data, the Operating System, type of hardware and software, and the configurations [20].

Multiple approaches exist for data recovery from hard drives, with disk imaging representing a significant method wherein bit-by-bit copies of disks are generated, even when physical damage is present. Disk imaging proves beneficial by restoring retrievable data from the disk while bypassing commands that could otherwise trigger a complete process restart upon error detection. It is best practice to conduct forensic analysis on the forensic copy of the evidence item, instead of directly examining the original, to avoid evidence contamination [41]. Forensic copies can be created through disk imaging, using specialist tools. This is an important step to ensure the integrity of the evidence item.

In a situation where a file system is damaged or deleted, retrieving a specific file becomes exceedingly challenging. To address this, file carving, also known as "carving," is a forensic technique utilized for data recovery [40]. File carving, a software technique crucial in digital forensics, has become increasingly important due to the proliferation of digital devices and storage types. File carving involves extracting structured data from raw data based on format-specific characteristics inherent in the structured data. It proves invaluable for recovering files and file fragments in scenarios where directory entries are corrupt or missing file metadata, particularly in forensic investigations involving criminal cases for evidence recovery. To effectively use file carving, knowledge of file headers and footers is necessary [18]. For instance, JPEG files have hexadecimal file header of FF D8, GIF has hexadecimal header of 47 49, EXE has a hexadecimal file header of 47 49, respectively.

A related study conducted a comparative analysis of file carving directly from the hard drive versus from a drive image, employing open-source tools like PhotoRec for file carving and disk imaging tools like gddrescue and Safecopy [3]. The research revealed a marked enhancement in data recovery efficiency when employing file carving from a disk image compared to directly from the HDD, concluding that file carving from a disk image represents a more effective and secure approach for recovering data from damaged or corrupted HDDs.

C. **Permanent Deletion of Data**

Data removal from computer disks could be quite challenging due to the lack of effective tools and misconceptions about data deletion [30]. While the format command shows graphically in the interface that a deletion or erasure has taken place, in truth, the data can be recovered. This poses a challenge in situations where there is a real need to erase data irreversibly. This is even more so today, with increasing awareness for data security and information governance [10]. In addressing this, the United States Department of Defense specified a standard (DoD 5220.22-M standard) in the National Industry Security Program Operating Manual, outlining the process to overwriting hard disk with ones and zeros. These involved three stages of overwriting passes verification [43].

The three stages are [22]:

Pass 1: Writes a zero and verify write

Pass 2: Write a one and verify write

Pass 3: Write a random character and verify write

In 2001, the Department of Defense issued an extended standard, providing for additional overwriting and verification method. The DoD 5220.22-M ECE method is a 7-Pass version of the DoD 5220.22-M, which runs the DoD 5220.22-M twice, with an extra pass (DoD 5220.22-M (C) Standard) in between [43]. However, despite this updated standard, DoD 5220.22-M standard remains the more popular, and widely accepted as the industry standard for secure and permanent erasure of data.

While DoD 5220.22-M remains a standard guide, however there are other sanitization standards, such NIST 800-88 Clear and NIST 800-88 Purge (Guidelines for Media Sanitization), which has showed remarkable advantages over the DoD standard. For instance, it is noted that the DoD 5220.22-M is more resource demanding and less economical than the other standards, and even the DoD has stopped referring to it [43].

The NIST Special Publication 800-88, “Guidelines for Media Sanitization” offers a methodical guideline for permanent erasure of data from electronic storage media. Originally released in 2006 as NIST 800-88, it was updated in 2014 as “NIST Special Publication 800-88 Rev. 1” (“NIST SP 800-88, Rev.1”). NIST 800-88 is

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

renowned for its three data sanitization categories of Clear, Purge, and Destroy. Unlike the DoD 5220.22-M, NIST 800-88 applies to all electronic storage media types [12]. While it was published for government use, the popularity has grown in both government and private sector.

D. Privacy and Data Protection laws

The proliferation of digital data has made data protection a critical issue globally. Several jurisdictions have come up with regulations and/or legislation to protect citizen data. Recent data privacy regulations like the EU General Deta Protection Regulation (EU-GDPR) and California Consumer Privacy AAct (CCPA) have introduced new requirements for timely and persistent deletion of user data [5]. For instance, the EU-GDPR is a landmark set of data protection rules, which became law on May 25th, 2018. It seeks to minimize the quantity of personal data collected by companies and enforce penalties for violation. GDPR increases the requirements for Data Privacy and Cyber Security operations in businesses and organizations [28]. Following this regulation, organizations that are in breach have been made to pay fines for non-compliance. Assessed fines are based on the organization's annual revenue. Such risk of hefty fines underscores the high premium on compliance and serves as a deterrent to erring institutions.

In Nigeria, the Nigeria Data Protection Regulation (NDPR) was released in 2019. The Nigerian Information Technology Development Agency (NITDA) [7] released the regulation. This regulation was pursuant to the NITDA Act 2007, which gave the organization the mandate to develop "guidelines for electronic governance". It is noteworthy to mention that NITDA's first attempt at releasing a data protection regulation in 2013 was fraught with challenges, and deficiencies in content, and was largely unenforceable [25]. The NDPR seeks to enforce the principle of data protection for data belonging to Nigerians both within and outside Nigeria; this includes data belonging to a suspect of a case.

The NDPR aligns with global data protection standards, such as the GDPR, emphasizing accountability, transparency, and the rights of data subjects. However, following greater calls to strengthen the enforcement mechanism, this resulted in the Nigeria Data Protection Act, passed in 2023 [39]. The Nigeria Data Protection Act of 2023 further strengthened the legal framework, addressing some gaps in the NDPR [2]. The NDPA serves as a legal framework to safeguard the personal information of Nigerian Data subjects. The Act was designed to protect the privacy rights of individuals by regulating the collection, storage, processing, and dissemination of personal data. The Act mandates that organizations handling personal data must implement appropriate security measures to prevent unauthorized access, alteration, disclosure, or destruction of personal data [33].

E. NDPA and Privacy Concerns with Refurbished Hard Drives

The Nigeria Data Protection Act [39] outlines specific obligations for data controllers and processors regarding data security and the disposal of personal data. Section 2.5.3 of the NDPR states that data controllers must ensure the secure disposal of personal data when it is no longer needed for the purpose it was collected. This includes implementing measures such as data anonymization, encryption, and secure data destruction methods [8]. In addition, Section 39 of the Nigeria Data Protection Act, 2023 also outlines the responsibilities of data controllers and data processors in implementing necessary technical and operational safeguards to ensure the security, confidentiality, and integrity of personal data in its possession.

Refurbished hard drives are often resold in secondary markets without adequate measures to ensure that all data from previous users are completely erased. In a bid to cut costs and perhaps as part of the organization's green policy, some organizations resell used computers and hard drives. While this may be well-intentioned, however, there is the risk of the computers and hard drives retaining sensitive corporate and personal information. Studies have shown that many refurbished drives sold online contain recoverable data, indicating insufficient data erasure practices [29]. This raises significant concerns about the effectiveness of existing data protection measures. Inadequate data-wiping practices can lead to the exposure of Personally Identifiable Information (PII), posing significant privacy risks to individuals. PII includes information such as names, addresses, social security numbers, and financial details, which can be exploited for identity theft, fraud, and other malicious activities. Forensic wiping involves the complete erasure of data from a storage device, making it unrecoverable even with advanced data recovery tools. This process is essential for protecting PII and other sensitive information when disposing or reselling hard drives.

F. Recent Studies and Recurring Issues

Data privacy and security breaches resulting from improper disposal or resale of used hard drives can have significant implications for individuals and organizations. One effect of ruminant information on hard drives is identity theft [13]. The effects of identity theft could range from taking bank loans to registering a social media account, using the victim's identification for the purchase of goods, which could be illegal, and collecting social benefits and medical insurance. In 2021, the Federal Trade Commission received 1.4 million reports of identity theft, making it one of the most common categories of consumer complaints [21]. The consequences of data theft

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

can be severe, ranging from financial loss and identity theft to reputational damage and legal penalties. As technology continues to advance, the importance of safeguarding data and combatting data theft remains a critical aspect of modern cybersecurity [47]. Since many individuals choose to donate - and there certainly is a need for these computers - it becomes even more imperative that individuals are made aware of the actions needed to completely erase their hard drives.

Kaspersky's Global Research and Analysis Team (GReAT) examined security in secondhand devices and found that 74% of used hard drives sold online contained residual data, including sensitive personal information. The study highlighted the significant risk posed by improper data sanitization practices, revealing that most secondhand devices still contained recoverable data, potentially exposing previous owners to data breaches and identity theft [35]. Another investigation by Blancco Technology Group revealed that 67% of the used hard disk drives and solid-state drives hold personally identifiable information and 11% contain sensitive corporate data [9].

The improper disposal of electronic devices poses a significant threat to data privacy and security, particularly in the healthcare sector. These breaches not only pose a financial and operational threat to hospitals but also present significant challenges to healthcare providers and clients. Numerous instances of data breaches have been attributed to this issue, exposing sensitive personal and medical information. There is a case of data breach in a healthcare institution in Waterville, Maine, where improper hard drive disposal potentially compromised close to 600,000 records [16]. Similarly, researchers reported a case where photocopiers used to process sensitive medical information were resold without data wiping, exposing patient data during storage [17]. This incident resulted in a $1,215,780 settlement with the U.S. Department of Health and Human Services. These incidents highlight the widespread nature of the problem.

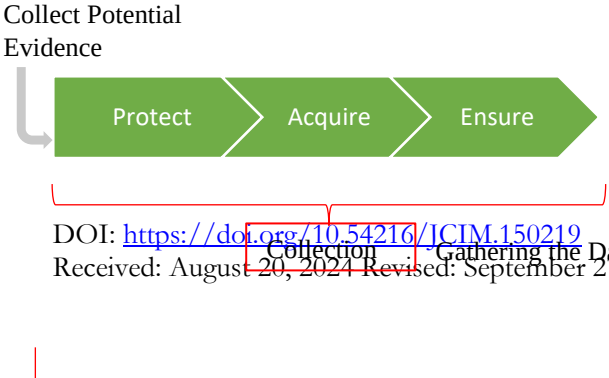
A recent study found that 70% of individuals in India are vulnerable to data breaches and privacy risks when disposing off used devices [42]. Similarly, investigations of second-hand storage devices in Thailand found that most contained personal identifiable information (PII) from previous owners [23]. The potential consequences of such leaks are dire, ranging from identity theft and financial fraud to reputational damage and legal repercussions for organizations. The ubiquity of electronic devices, coupled with the increasing volume of sensitive data stored on them, necessitates a reevaluation of disposal practices. Secure data deletion protocols and responsible recycling initiatives are crucial to mitigating the risks associated with improper electronic waste management.

Several recent research have shown that the issue of the challenge of residual data in secondhand storage device [24]. For instance, in another study, the researchers procured second-hand memory cards from the Australian eBay site, and the results indicated that resold memory cards were disposed of insecurely, with personal or business confidential data either undeleted or easily recoverable [44]. In another study involving researchers from Avast, 20 Android devices were procured from eBay, and the researchers found considerable amounts of sensitive data on them including pictures, contacts, chat logs, search history, and location history, among others [6]. Another study found 36,136 recoverable files including a range of data detailing private information of previous owners, and confidential corporate data, with 20% of the purchased USB devices securely wiped before sale. This research showed that Personally Identifiable Information such as user credentials, product recipes, and customer credit card data could be recovered from the employees' USB flash drives [46]. This could potentially have a high-risk impact on the company, such as reputational damage and sabotage of products by competitors.

3. Materials and Methods

The primary purpose of this research is to establish the presence of residual data and Personally Identifiable Information on refurbished hard drives due to improper sanitization. Understanding the complexities of data recovery is critical in a digital landscape dominated by consumer-grade storage devices. This section presents the methodology, tools, and considerations required to successfully recover erased/deleted data from routinely used off-the-shelf drives.

There is not a single universally accepted digital investigation framework or model out there. However, there are consistent steps that appear common to most. For this study, we shall adopt the steps in Digital Investigation outlined by NIST IR 8354 [32]. The NIST IR 8354 is reproduced in figure 1 below. For this study, this was slightly modified to reflect the realities of the study (hard drives were procured from a wholesale shop) and placed emphasis on specifics of the preparation stage. The modified model is shown in figure 2, showing the addition of the preparation stage.



DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

Figure 1. Steps in a digital investigation (NIST IR 8354, 2022).

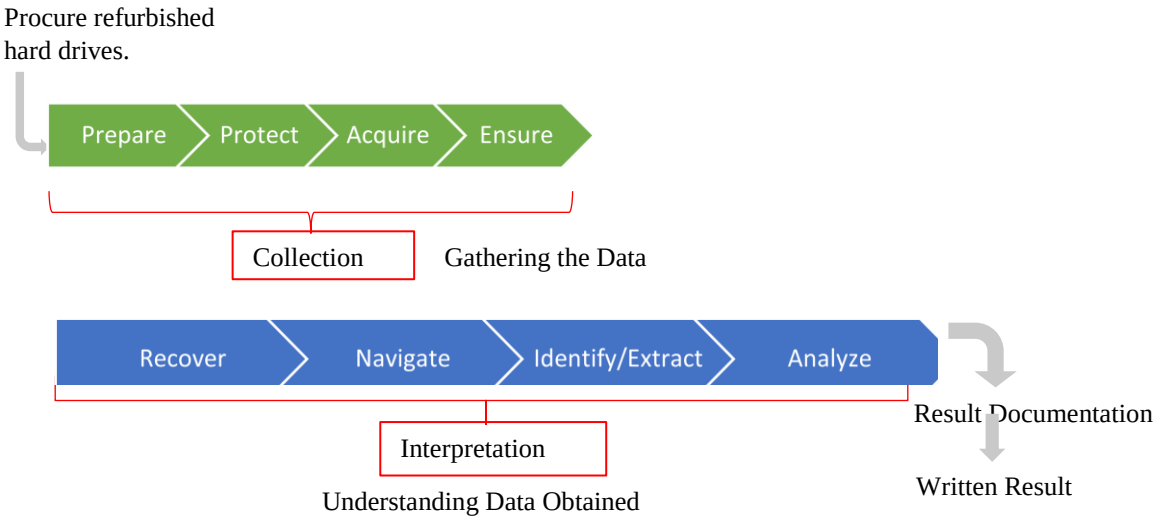


Figure 2. Steps in adopted in this study (Modified NIST IR 8354)

A. Procurement of Refurbished Hard Drives

This study investigated the presence of residual data, and very importantly Personally Identifiable Information, in refurbished hard drives sold in the market. For this study, the researchers bought 5 hard drives (500 GB each), randomly selected from a popular retail computer and accessories shop in Abuja, Nigeria. Three of these hard drives were examined and analyzed for the presence of residual data and PII. The other two were sanitized forensically and used to store forensic copies of the evidence on hard drives. In line with industry best practice, the forensic copies (and not the original copies) were analyzed.

B. Prepare

Given the significance of the "preparation" phase in a forensic investigation, this research incorporated it as a distinct step. Digital forensics is a meticulous discipline that demands careful planning and precision. During this stage, the researchers ensured that all necessary tools were available in the laboratory. The data recovery exercise was conducted at the Digital Forensic Laboratory of Digital Footprints Nigeria Limited, Abuja, Nigeria. Additionally, the researchers verified the validity of the software licenses to ensure they were up to date. They also confirmed with the laboratory manager when the tools were last independently validated. Documentation indicated that the tools had been validated eight days prior to the study's commencement. Further records revealed that the forensic tools were routinely validated every three months.

In the preparation stage, the two hard drives meant to store the forensic copies of the hard drive were forensically sanitized to ensure any data on it was permanently erased. Using Media Clone Super imager, this study adopted the DoD 5220.22-M standard also known as the National Industrial Security Program Operating Manual (NISPOM) [15], for data sanitization of HDD-1 and HDD-2 (the analysis hard drives). This standard ensured drives were over-written 3 times to permanently erase any residual data. Images were taken of these processes for proper documentation. Table 1 below shows information about the software and hardware tools used.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

Table 1: Software and hardware tools used.

S/N	Type	Name	Description
1.	Software	MediaClone SuperImager	It is a top-performance Field Computer Forensic Imaging tool and a Complete Digital Forensic Investigation platform. Allows for full forensic analysis, data extraction, and triage data collection. It also performs hard drive sanitization and secure wiping, using different standards.
2.	Software	EaseUS Data Recovery Tool	EaseUS data recovery software helps users to recover Documents, Graphics, Video, Audio, Email, and Other Files. It offers solutions for recovery of lost files from hard drive, disk, partition and other storage devices.
3.	Hardware	External Hard Drives (Seagate)	Five 500GB external hard drive were procured from retail computer shop
4.	Hardware	FRED-L Forensic Laptop	A FRED-L is a high end digital forensic laptop used for analyzing the evidence item.
5.	Hardware	Anti-static Mat	The hard drives were placed on the anti-static mat as a protective measure against anti-static discharge.
6.	Software	Autopsy	Autopsy is a digital forensics tool used for investigating and analyzing computer systems and digital storage devices.
7.	Hardware	Tableau Forensic USB 3.0 Bridge	This is a write blocker that was used to protect the integrity of the hard drives.

C. Protect

In line with best practices for conducting a forensic investigation, the researchers took steps to protect the integrity of the hard drives under investigation. To achieve this, the functionality of the write-protect tool on the Media Clone Super Imager was confirmed, ensuring that evidence remained uncontaminated, unmodified, and undamaged. This tool safeguards the drive by preventing any data from being written to or deleted from it. The researchers also used the Tableau Forensic USB 3.0 Bridge, a write-blocker that protects data from being written to the hard drives. Additionally, the hard drives were securely placed on an anti-static mat to protect them from electrostatic discharge.

D. Acquire

In digital forensics, minimizing the direct handling of original evidence is a standard best practice. To adhere to this, the researchers forensically duplicated the three hard drives purchased; ensuring that the analysis was conducted on forensic images rather than the original drives [18]. The Media Clone Super Imager was utilized for this duplication process, enabling the analysis to focus exclusively on the forensic copies. Before duplication, as outlined in the preparatory phase, the destination hard drives were forensically sanitized to ensure they were free of any residual data. Subsequently, forensic copies of the three hard drives were created and securely stored on the sanitized drives.

E. Ensure

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

The researchers ensured the integrity of the hard drives. To achieve this, the forensic duplication process created an MD5 AND SHA256 hash of both source and destination hard drives to verify the integrity of the data. Hashing is a technique of collecting a group of data and compressing it into a certain length of alphanumeric characters, which is usually shorter than the original data [11]. It uniquely identifies a file or drive in its current state, such that if anything changes, the hash will change. This is a scientific way of determining the integrity of the drive. To ensure the research conforms to digital forensic process, all investigative details and steps were recorded in the chain of custody form.

F. Recover

For this study, the EaseUS Data Recovery Wizard was employed, as the research focuses on a software-based recovery method. EaseUS Data Recovery Wizard is a widely recognized tool for data recovery, extensively tested and validated in both industry applications and scholarly research [36]. In line with digital forensics best practices, all preparatory steps taken in this study were documented.

G. Navigate

After recovering data, the next step is to navigate and examine it [32]. This process is primarily carried out using specialized data recovery tools. In this study, EaseUS Data Recovery Wizard and Autopsy were employed to effectively parse the recovered data structures and associated metadata. These tools ensured that the data was navigable and accurately displayed in its original context, facilitating a thorough examination of the recovered information.

H. Identify/Extract

This step involves identifying and extracting data of interest. In this study, there were two levels of interest. First was to generally look out for residual data. To achieve this, the study aggregated all the data recovered by the EaseUS Data Recovery tool. The next level was to look out for PII from the recovered data. Since this study did not have any background knowledge of the previous owner of the drives, there are no specific keywords known. However, since our target is PII, there are wide possibilities, such as searching for all picture files, documents, and spreadsheets, semblance of credit cards, email addresses, phone numbers, and text files. For a focused search, the forensic image was loaded on Autopsy tool. Autopsy is an opensource tool that is widely respected [1, 37]. The tool automatically groups similar data types and flags suspicious files as well as PII such as credit card, phone numbers, EXIF metadata, and others.

I. Analyze

The recovered data were further analyzed and classified in terms of data types. Personally, Identifiable Information were spotted and flagged. In deference to the privacy of the affected, all PII were redacted in this paper. The researchers leveraged the analytical capability of both EaseUS Recovery Wizard and Autopsy forensic tool to analyze and recreate events.

J. Result Documentation

The documentation stage, following the comprehensive analysis of the data recovery process, served as the conclusive phase of the study. This stage in the methodology involves meticulous record keeping of every step, configuration setting, and observation during data recovery. This ensures reproducibility and reliability of results, contributing to the credibility of research findings and forensic practices, and serves as a valuable resource for subsequent analyses.

4. Result and Findings

This session presents the results of this study. Five hard drives were procured from a retail computer shop. Two (HDD-A1 and HDD-A2) were forensically sanitized, to serve as the analysis drive. The other three hard drives (HDD-1, HDD-2, and HDD-3 were forensically imaged, with HDD-A1 as destination for the forensic image of HDD-1 and HDD-2, while HDD-A2 served as the destination for the forensic image of HDD-3. Table 2 below shows the time taken for wiping and cloning the hard drives.

Table 1: Hard Drives and Summary of Processing Time

Hard-drive	Storage Capacity	Description	Duration
HDD-A1	500GB	Hard drive was forensically sanitized and used as destination for forensic image of HDD-3 and HDD-4, respectively, following the DoD 5220.22-M standard.	4hrs

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

HDD-A2	500GB	Hard drive was forensically sanitized and used as destination for forensic image of HDD-5, following the DoD 5220.22-M standard.	4hrs
HDD-1	500GB	Original hard drive that was cloned	2hrs
HDD-2	500GB	Original hard drive that was cloned	2hrs
HDD-3	500GB	Original hard drive that was cloned	2hrs

As seen in the table 2 above, there is a huge time difference between wiping the first two hard drives and mirroring the other three hard drives. This is because the first two hard drives were over written following DoD 5220.22-M standard which is using at least 3 iterations to completely erase data from the hard drives, using the super imager. Table 3 below shows the specifications of the system used for analysis.

Table 2: System specifications

Forensic Laptop	FRED-L
Processor	AMDD Ryzen 7 3700X 8-Core Processor 3.59GHz
Operating System	Windows 10 Pro
RAM	64GB
Software Used	EaseUS Data Recovery Wizard (v17.0.0.0)

A. **Sanitization of the Hard Drives**

HDD-A1 and HDD-A2 were forensically wiped using Super Imager, a renowned tool in digital forensics. The sanitization process followed the DoD 5220.22-M standard. This three-iteration process ensures irreversible data removal, setting the stage for a detailed examination of the data recovery methodology and its efficacy.

B. **Forensic Imaging of the Hard Drives**

A forensic image of HDD-1, 2, 3 was made using the Super Imager. This process ensured a bit-by-bit copy of the data from the source to the target drives and generated MD5 and SHA256 hash values, acting as digital fingerprints, verifying data integrity and authenticity, and reinforcing forensic procedures.

C. **Hash Verification of the Hard Drives**

Figure 3, 4, and 5 below show the MD5 and SHA256 hash verification of the three hard drives respectively. The hash of both “source” and “target” hard drives verified the integrity of the forensic images.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

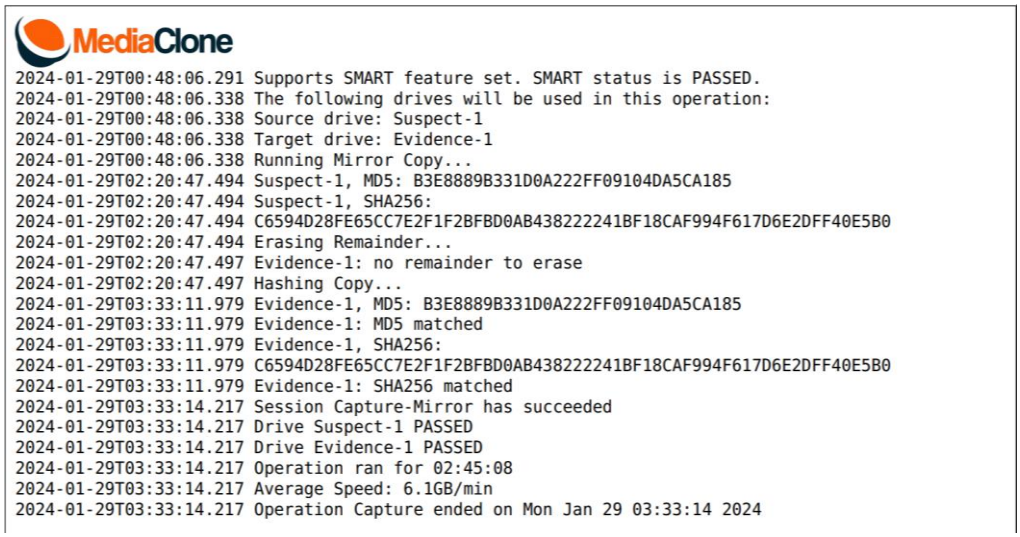


Figure 3. Imaging showing the hash verification of HDD-3

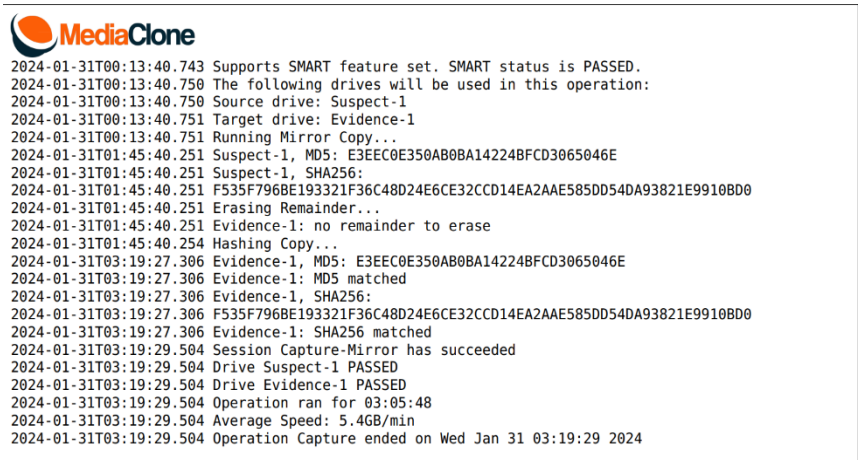


Figure 4. Imaging showing the hash verification of HDD-4

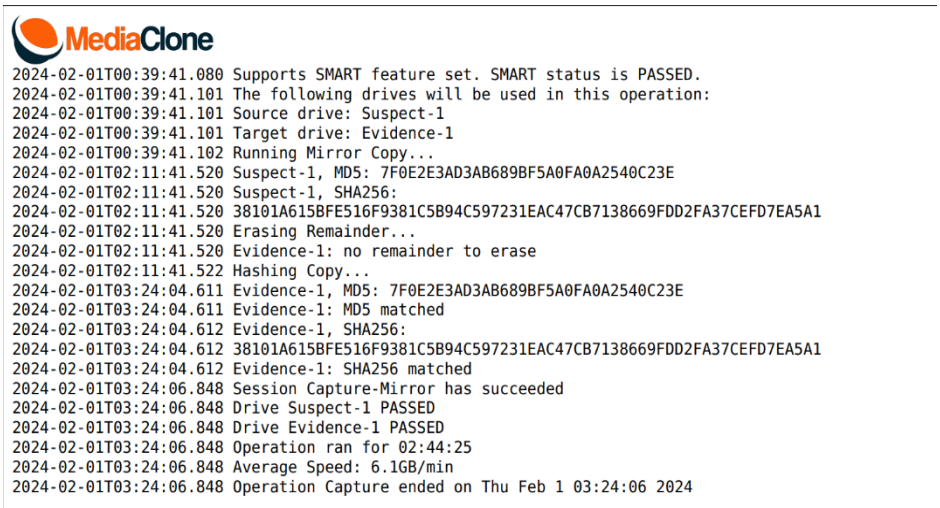


Figure 5. Imaging showing the hash verification of HDD-5

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

D. Analysis/Findings

The data recovery process using EaseUS Data Recovery Wizard demonstrated notable success, with HDD-1 yielding substantial results as the tool effectively recovered 28,691 files containing 152.20 GB worth of data encompassing a diverse array of files, including Personally Identifiable Information (PII) such as pictures, sensitive documents, videos, and audio files. HDD-2 and HDD-3 did not contain any meaningful residual data or PII. Specifically, 5 files containing 32.27KB worth of data were recovered from HDD-2, while 0 files worth 0KB were from HDD-3. This outcome aligns closely with the research objectives, showing the presence of residual data and PII in refurbished hard drives bought for this study. The successful recovery from this HDD-1 also underpins the effectiveness of the chosen data recovery methodology, highlighting the tool's ability to recover valuable information crucial to this research.

Analysis of the recovered files from HDD-1 shows the drive mostly had pictures, suggesting that the owner may have used it for storage of personal and professional photographs. Specifically, 21,627 files, representing 78% percent of all files recovered, were pictures. Most of the recovered pictures are portraits and photographs of humans that can be easily identifiable given the clarity and quality. Quite disturbingly are many sensitive and personal pictures. For instance, more than 400 photographs of schoolchildren in their official school uniforms were recovered; however, the researchers excluded these images from the paper to protect privacy. Such photographs, if obtained by malicious actors, could be exploited for criminal activities, including cybercrime, identity theft, and even facilitating kidnappings or other harmful schemes.

Table 4 and Figure 6 below shows details (showing file count and file type) of the data recovered from the HDD-1. Figure 6 is a pie chart showing the different file types recovered.

Table 4: Summary of recovered data from HDD-1 on EaseUS

S/N	Files Found	Count
1.	Pictures	21627
2.	Videos	199
3.	Documents	1607
4.	Archives	2147
5.	Audio	2061

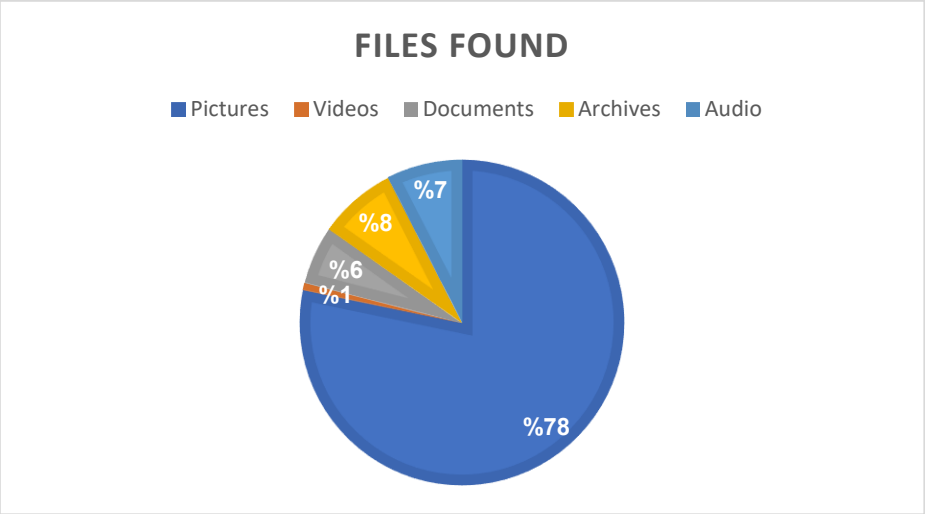


Figure 6. Chart showing the files found from HDD-1.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

Autopsy forensic tool was used to cross-validate the data recovery process on HDD-1 after initially using EaseUs to recover a significant number of files. Autopsy analysis revealed a substantial number of artifacts, including deleted files and orphan files. Orphan files are files that have lost their directory references and are no longer accessible through the file system's directory structure, typically due to deletion or corruption. The detailed analysis confirmed the presence of 118,035 orphan files, providing critical insights into the data remnants on the hard drives. Table 5 (count by file type) and 6 (count by recovery location) shows the data recovered using Autopsy tool. Analyses of recovered files shows several PII, including pictures of schoolchildren and other sensitive pictures, which for privacy reasons are not included in this paper. Further analysis of the metadata of the recovered pictures showed that 2872 of the pictures were captured with a Canon Camera, while NIKON D610 camera accounted for 2113 pictures. Sony and Apple devices both accounted for 1 picture each.

Table 5: Summary of recovered data (count) from HDD-1 on Autopsy

S/N	File Found	Count
1.	Images	4990
2.	Videos	1
3.	Audios	3
4.	Documents	490
5.	Executables	347
6.	Unknown	3992
7.	Other	156
8.	Not Analyzed	88,491

Table 6: Summary of recovered data (location) from HDD-1 on Autopsy

S/N	File Types	Count
1.	Allocated File	430
2.	Unallocated Files	98,040
3.	Slack Files	6
4.	Directories	30,131

The EaseUs tool recovered five ‘lost files’ from HDD-2. These files are suspected to be deleted files. In the lost file directory, there are 2 folders named DIR3 and DIR4 files with the .bup and .dat file extensions. The .bup extension is commonly associated with backup files, protecting critical data from loss or corruption, while the .dat extension refers to generic data files, whose interpretation depends on the application.

In the analysis of HDD-2 using Autopsy, 7 files were recovered from $Extend directory, with 2 of these files located in the $Deleted subdirectory. Additionally, the hard drive had 8 files in the $RmMetadata directory, 3 files in the $Recycle.Bin directory, and 466 unallocated files. The metadata for the deleted files shows they were last modified, accessed, and created on the same date, indicating they may have been systematically managed or deleted around the same time. Despite the presence of these files, the overall volume of recoverable data on this drive is significantly lower compared to HDD-1, suggesting either a more thorough data cleaning process or less overall usage of the drive.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

The data recovery process using EaseUS Data Recovery Wizard on HDD-3 yielded no recoverable files. Despite diligent efforts, the absence of any retrieved data prompts a closer examination of potential factors contributing to this outcome. Autopsy result for HDD-3 had 0 Orphan Files, 1 Carved file, $Extend directory contained 7 files, and 466 unallocated spaces. Table 7 summarises the results from thee three hard drives analysed in the study.

Table 7: Summary of results from all hard drive

Hard Drive	Number of Files Recovered	Volume of Data	Key Findings
HDD-1	28,691 files	152.20 GB	Majority were pictures (21,627), sensitive pictures including images of school children, highlighting a failure in sanitization.
HDD-2	5 files	32.27 KB	Residual files included .bup and .dat extensions; sanitization was incomplete but improved compared to HDD-1.
HDD-3	0 files	0 KB	No data recovered, demonstrating effective sanitization.

5. Conclusion

This study was done to show the presence of residual data and particularly Personally Identifiable Information (PII) from refurbished hard drives sold in the open market. The study explored data recovery from refurbished hard drives, highlighting the complexity of the process and the need for a balance between technology, ethics, and legal standards. Results from this study showed the presence of residual data, with one out of the three hard drives containing 28,691 files, with total size of 152.20GB. These recoveries contained Personally Identifiable Information (PII) such as pictures, documents, videos, and audio files. Specifically, there were more than 400 pictures of schoolchildren in their school uniforms. The study identified the threat of poor sanitization protocol and practice in organizations giving out their computers and hard drives for resale or as gifts.

The various legislations around the world, such as the GDPR, NDPR, and others, all highlight the responsibility of data processors and administrators and the need for safe disposal of computers and hard drives. This study showed that despite these legislations, many users are not aware of the technical responsibilities. The findings of this study necessitate the need for a recommendation for a collective review of data recovery techniques, asking stakeholders to prioritize the security of digital information.

The study showed that the sale of refurbished hard drives containing residual data presents significant information governance challenges. From a legal perspective, the unauthorized retention and distribution of personal or sensitive information on these drives may violate data protection laws, leading to potential legal consequences for the sellers. For instance, several PII, including sensitive pictures of children and others are a clear breach of privacy, and the laws such as GDPR and NDPA. However, with breaches like this, this study shows much needs to be done to drive data protection awareness amongst stakeholders. For instance, stakeholders need to be properly informed and made aware of safe deletion and sanitization of hard drives before reselling or giving it out as gift.

In addition, with such brazen privacy violations such as this, this study points out the need for strengthening of enforcement mechanisms. Ensuring compliance with data protection regulations and instituting robust data sanitization practices are imperative to uphold the trust and ethical responsibility associated with the sale of refurbished or second hand hard drives. This study showed the need for industry-wide standardization and transparency in the refurbishing process, as residual data on off-the-shelf drives highlights the need for responsible practices.

Acknowledgment

This research was conducted at the Laboratory of Digital Footprints Nig. Limited, Abuja, Nigeria. The authors are grateful to management of Digital Footprints Nigeria Ltd for all the technical support.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

References

[1] Adamu, H., Ahmad, A. A., Hassan, A., & Gambasha, S. B. (2021). Web Browser Forensic Tools: Autopsy, BHE and Net Analysis. International Journal of Research and Innovation in Applied Science, 06(05), 103–107. https://doi.org/10.51584/ijrias.2021.650

[2] Adeoti, E. (2023, July 24). A New Era of Data Protection and Privacy; Unveiling Innovations & Identifying Gaps in the Nigeria Data Protection Act of 2023. Social Science Research Network. https://doi.org/10.2139/ssrn.4520238

[3] Aljumah, A., Uddin, M. Y., & Ahamad, M. G. (2014, December). Comparison between file carving from disk drive and disk image in open-source environment. In International Conference on Computing and Communication Technologies (pp. 1-4). IEEE.

[4] Angamutu, K. A., Rahman, N. A. A., & Suki, N. N. A. N. (2020). A Customized Data Recovery Tool. Journal of Physics: Conference Series, 1712(1), 012019. https://doi.org/10.1088/1742-6596/1712/1/012019

[5] Athanassoulis, M., Sarkar, S., Papon, T.I., Zhu, Z., & Staratzis, D. (2022). Building Deletion-Compliant Data Systems. IEEE Data Eng. Bull., 45, 21-36.

[6] Avast. (2014, July 8). Tens of thousands of Americans sell themselves online every day. Blog.avast.com. https://blog.avast.com/2014/07/08/tens-of-thousands-of-americans-sell-themselves-online-every-day/

[7] Babalola, O. (2021). A bird's eye rundown on Nigeria's Data Protection Legal and Institutional Model. Gravitas Review of Business & Property Law, 1. 12(2). https://doi.org/10.2139/ssrn.4625918

[8] Banjo, A. A. (2020). The actualisation of personal data protection in Nigerian law: an analysis of personal data protection in the Nigerian and European Union legal systems. Dspace.ut.ee. https://dspace.ut.ee/items/ccd6f0c3-6503-4911-b2ef-e145cc517003

[9] Blancco Technology Group. (2016, August 18). - The Leftovers: A Data Recovery Study. https://www.ultimatewindowssecurity.com/blog/default.aspx/1000?p=5f833578-27a6-4a78-b5c5-671847aa0a77

[10] Casey, E. (2011). Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet (3rd ed.). Academic Press.

[11] Chi, L., & Zhu, X. (2017). Hashing techniques: A survey and taxonomy. ACM Computing Surveys (Csur), 50(1), 1-36.

[12] Cullipher, V. (2019, May 9). What is NIST 800-88, and What Does "Media Sanitization" Really Mean? · Blancco. Blancco. https://www.blancco.com/resources/blog-what-is-nist-800-88-media-sanitization/

[13] Dawn Medlin, B., Cazier, J. A., & Weaver, R. M. (2008). Consumer is PCs: A Study of Hard Drive Forensics, Data Recovery, and Exploitation. Journal of Information Privacy and Security, 4(3), 3–15. https://doi.org/10.1080/2333696x.2008.10855843

[14] Dayma, H., & A, R. (2024). Rekindling Digital Remnants: A Comprehensive Exploration of Data Restoration Pathways. International Research Journal of Modernization in Engineering Technology and Science, 06(05). https://doi.org/10.56726/irjmets56582

[15] Department of Defense. (2020, December 21). National Industrial Security Program Operating Manual (NISPOM). Unblock.federalregister.gov. https://www.federalregister.gov/documents/2020/12/21/2020-27698/national-industrial-security-program-operating-manual-nispom

[16] Dillard, G. (2022, April 25). Column: Improper disposal of hard drives can lead to health records breaches. The Business Journal. https://thebusinessjournal.com/column-improper-disposal-of-hard-drives-can-lead-to-health-records-breaches/

[17] Dort, K. K., & Capizzi, M. D. (2013, September 8). Photocopiers – A Recurring Data Security Risk | Publications | Insights | Faegre Drinker Biddle & Reath LLP. Www.faegredrinker.com. https://www.faegredrinker.com/en/insights/publications/2013/9/photocopiers--a-recurring-data-security-risk

[18] Easttom, C. (2019). System forensics, investigation, and response. Jones & Bartlett Learning.

[19] Engin Z. & Arslan, S. S. (2020). Cloud^{2}HDD: Large-Scale HDD Data Analysis on Cloud for Cloud Datacenters. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.1109/icin48450.2020.9059482

[20] Eoghan, C. (2000). Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet | Office of Justice Programs. Www.ojp.gov. https://www.ojp.gov/ncjrs/virtual-library/abstracts/digital-evidence-and-computer-crime-forensic-science-computers-and

[21] Federal Trade Commission. (2022, February 22). New Data Shows FTC received 2.8 million Fraud Reports from Consumers in 2021. Federal Trade Commission. https://www.ftc.gov/news-

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

events/news/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021-0

[22] Fisher, T. (2023, September 20). Data Sanitization Methods: Everything You Need to Know. Lifewire. https://www.lifewire.com/data-sanitization-methods-2626133

[23] Funge-Smith, M., & Beokhaimook, C. (2023). Investigation and Analysis of Information Remaining on Used HDDs in Thailand. 2023 8th International Conference on Business and Industrial Research (ICBIR), 306-311.

[24] Garfinkel, S. L., & Shelat, A. (2003). IEEE Security & Privacy: Data Forensics - Rememberance of Data Passed: A Study of Disk Sanitization Practices. IEEE Distributed Systems Online, 4.

[25] Greenleaf, G. (2019). Nigeria Regulates Data Privacy: African and Global Significance (pp. 23–25). (2019) 158 Privacy Laws & Business International Report. https://ssrn.com/abstract=3401783

[26] Hepisuthar, M. (2021). Comparative analysis study on SSD, HDD, and SSHD. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(3), 3635-3641

[27] Howarth, J. (2022, July 1). 30+ Identity Theft Statistics for 2023. Exploding Topics. https://explodingtopics.com/blog/identity-theft-stats

[28] Imrichová, A. (2020). GDPR impact on Information Security Incident detection and response. https://is.vsfs.cz/th/rx5w3/Andrea-Imrichova_-_GDPR_Impact_on_SOC.pdf

[29] Jones, A., Valli, C., Sutherland, I., & Thomas, P. (2006). The 2006 Analysis of Information Remaining on Disks Offered for Sale on the Second-Hand Market. Journal of Digital Forensics, Security and Law, 1(3). https://doi.org/10.15394/jdfsl.2006.1008

[30] Jones, A., & Afrifa, I. (2020). An Evaluation of Data Erasing Tools. The Journal of Digital Forensics, Security and Law. https://doi.org/10.15394/jdfsl.2020.1615

[31] Kissel, R., Regenscheid, A., Scholl, M., & Stine, K. (2014). Guidelines for Media Sanitization. Guidelines for Media Sanitization, 1. https://doi.org/10.6028/nist.sp.800-88r1

[32] Lyle, J. R., Guttman, B., Butler, J. M., Sauerwein, K., Reed, C., & Lloyd, C. E. (2022). Digital Investigation Techniques: Digital investigation techniques: a nist scientific foundation review. https://doi.org/10.6028/nist.ir.8354

[33] Nigeria Data Protection Act. (2023). the Federal Government Printer, Lagos, Nigeria. https://ndpc.gov.ng/Files/Nigeria_Data_Protection_Act_2023.pdf

[34] Pecht, M., & Elburn, E. (2020). Commercial hard drive failures in a data center application and the role of SMART attribute information. Circuit World, ahead-of-print (ahead-of-print). https://doi.org/10.1108/cw-07-2020-0127

[35] Pike, S. (2021, January 29). Uncovering private data in secondhand sales. https://www.kaspersky.com/blog/data-on-used-devices/38610/

[36] Putra, A., Siahaan, M. D. L., & Arpan, A. (2022). Comparative analysis of data recovery using easeus data recovery wizard and recuva applications. Infokum, 10(03), 161–165. https://infor.seaninstitute.org/index.php/infokum/article/view/686

[37] Sachdeva, S., B.L., R., & Sharma, A. (2020). Analysis of Digital Forensic Tools. Journal of Computational and Theoretical Nanoscience, 17(6), 2460–2468.

[38] Sibe, R.T. (2022). Africa's Chaotic Legal and Regulatory Cybersecurity Landscape Requires Harmonization. Forbes Technology Council. https://www.forbes.com/sites/forbestechcouncil/2022/08/02/africas-chaotic-legal-and-regulatory-cybersecurity-landscape-requires-harmonization/

[39] Sibe, R. T., & Kaunert, C. (2024). Conclusion and Recommendations for Digital Forensic Readiness of Nigerian Financial Crimes Agencies. In Cybercrime, Digital Forensic Readiness, and Financial Crime Investigation in Nigeria (pp. 179-207). Cham: Springer Nature Switzerland.

[40] Sibe, R. T., & Bekom, D. (2025). Digital Forensic Investigation of an Unmanned Aerial Vehicle (UAV): A Technical Case Study of a DJI Phantom III Professional Drone. Journal of Cybersecurity and Information Management (JCIM) Vol, 15(01), 197-210.

[41] Singh, A., Kumar, S., & Singh, V. (2020). Extraction and analysis of forensic deleted datafrom digital evidence using the sleuthkit. International journal of multidisciplinary educational research, 9(10(7)).

[42] Stellar. (2019). Residual data study on second hand devices: a study on the risk implication for people, businesses and economies. https://www.stellarinfo.com/pdf/Stellar-Residual-Data-Study-on-Second-Hand-Devices-Report-April-2019.pdf

[43] Stiennon, R. (2019, March 28). Everything You Need to Know About the DoD 5220.22-M Wiping Standard & Its Applications Today. Blancco Technology Group. https://www.blancco.com/blog-dod-5220-22-m-wiping-standard-method/

[44] Szewczyk, P., Sansurooah, K., & Williams, P. A. H. (2018). An Australian longitudinal study into remnant data recovered from second-hand memory cards. International Journal of Information Security and Privacy, 12(4), 82–97. https://doi.org/10.4018/ijisp.2018100106

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024

[45] Varayogula, S. N., Dodiya, K., Lakhalani, P., & Chawla, A. (2022). Computer forensics data recovery software: A comparative study. International Journal of Innovative Research in Computer Science & Technology, 10(2), 513–518. https://acspublisher.com/journals/index.php/ijircst/article/view/10623

[46] Widya Chaerani, Clarke, N., & Bolan, C. (2011). Information leakage through second hand USB flash drives within the United Kingdom. https://doi.org/10.4225/75/57b2ba7e40cea

[47] Xin,L.T., M. I. Dulloo, M. H. Majeed, J. P. H. Wan, H. Azam, and S. R. Sindiramutty, "Cybercrime Unmasked: Investigating Cases and Digital Evidence," International Journal of Emerging Multidisciplinaries: Computer Science & Artificial Intelligence, vol. 2, no. 1, Nov. 2023, doi: https://doi.org/10.54938/ijemdcsai.2023.02.1.255.

[48] Zeydan, E., & Arslan, S. S. (2020, February). Cloud 2 HDD: large-scale HDD data analysis on cloud for cloud datacenters. In 2020 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN) (pp. 243-249). IEEE.

DOI: https://doi.org/10.54216/JCIM.150219
Received: August 20, 2024 Revised: September 27, 2024 Accepted: November 19, 2024