Digital Forensics applications towards digitized collections in Cloud

*a process approach to gathering evidences for authenticity, integrity and accessibility*

Sanjay Singh

Digital Curation, masters level (120 credits)
2017

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering
Digital Forensics applications towards digitized collections in Cloud: a process approach to gathering evidences for authenticity, integrity and accessibility

For the degree of MSc. in Digital Curation
Lulea University of Technology

Submitted by
Sanjay Singh

Department of Computer Science, Electrical and Space Engineering
June 2017
Declaration

I declare that the current Master thesis with a title of “Digital Forensics applications towards digitized collections in Cloud: a process approach to gathering evidences for authenticity, integrity and accessibility” is submitted in the partial fulfilment of MSc (Digital Curation) programme at Lulea University of Technology, Sweden.

This thesis work is the result of my own independent research effort. It has not been submitted to any other institutions for any award. Wherever the thesis is indebted to the work of others, due acknowledgements have been made.

Sanjay Singh

Signature: ..................

Date: …02.June 2017……
Abstract

The growth of data/information on social media and in large organizations is huge in terms of velocity, volume and variety which is also something being tackled by the large IT companies providing Big Data solutions. The other challenges which are linked to managing the huge pile of data are about ensuring preservation and access of crucial data which has implications in every sector ranging from pharmaceuticals to aerospace and cultural institutions (museums, archives and governmental records).

The challenges for data management are further complicated by the changing infrastructure landscape and the new business models to host data in virtualized cloud-based storage termed as Cloud solutions (PaaS, SaaS, and IaaS). Several large companies and public institutions are migrating their data/applications to cloud due to the apparent benefits of scalability, reliability, cost, easy of operability and security.

The digitization and maintenance of e-records / digital archives in Cloud provides many potential benefits but it is also prone to several risks to ensure long-term retention of data as well as to ensure integrity, authenticity and accessibility of data. For several organizations such as memory institutions, heavy industries (Aerospace & Defence), banks and pharmaceutical companies, it is business critical to securely store data for long-term with integrity, authenticity and accessibility ensured. Hence, along with preservation of data, it is crucial to keep integrity and authenticity of data intact.

The digital forensics methods and tools offer several solutions to ensure preservation of data and detect risks at pre-ingest stage of digital archiving to take appropriate measures towards ensuring authenticity, integrity and accessibility. The specific forensics methods and tools also offers possibilities to detect malicious activities or tampering in the digital archives and prepare report for presentation in the court.

This thesis work is focussed on the applications of digital forensics towards ensuring the preservation of data in cloud-based storage. It discusses the applications of processes, methods and tools to improve the acquisition, management and accessibility of collections hosted on cloud-based storage (Google Drive, Sky Drive). The pilot platform (i.e. Google Drive) would be tested with forensics methods/tools to draw conclusions for the memory institutions about hosting their data on cloud storage.

Keywords: cloud, forensics, archiving, preservation.
Acknowledgements

I am indebted to many people who guided or supported me to complete my studies at Lulea University of Technology. Firstly, I would like to express my sincere gratitude to the thesis supervisor Jörgen Nilsson for his throughout guidance and support during the study programme. It was only because of his immense cooperation and support, I could finally complete my MSc Digital Curation programme.

I am also thankful to the course tutors Mari Runardotter and Hugo Quisbert who taught different modules in the course which had provided the required knowledge and toolset to develop ideas for this research work.

I am deeply thankful to EC Council, USA for developing excellent cyber security/forensics certification (CHFI) which had helped develop necessary technical background to explore the topic of Digital forensics in context of Digital archiving.

I am also grateful to Mr. Harshad Shah, Chief Information Security Officer, GCSRT (Global Cyber Security Incident Response Team) for helping to understand the criticalities and complexities of Cyber Security and Digital Forensics.

My special thanks to the industry colleagues, Mr. S.P. Singh (Senior Security Specialist, Phillips) and Mr. A.K Singh (GM, Enterprise IT, INTAS Pharma) who had helped me to shape the research idea with their contributions and valuable feedback.

Finally, I would like to thank my friends, family and colleagues who had helped and encouraged me to complete my research work.

Thank You!!
Sanjay Singh
# Table of Contents

ABSTRACT ................................................................................................................................. III

ACKNOWLEDGEMENTS ............................................................................................................. IV

LIST OF FIGURES ..................................................................................................................... 7

LIST OF TABLES .......................................................................................................................... 8

LIST OF ABBREVIATIONS .......................................................................................................... 9

CHAPTER 1 ................................................................................................................................. 10

INTRODUCTION ............................................................................................................................ 10

1.1. BACKGROUND OF THE STUDY ....................................................................................... 10

1.2. PROBLEM STATEMENT ..................................................................................................... 11

1.3. PURPOSE OF THE STUDY ................................................................................................. 12

1.4. RESEARCH QUESTION ....................................................................................................... 13

1.5. RESEARCH STRATEGY ....................................................................................................... 13

1.6. RELEVANCE OF STUDY ................................................................................................... 14

1.7. LIMITATIONS AND DELIMITATIONS ............................................................................... 14

1.8. THESIS STRUCTURE .......................................................................................................... 14

CHAPTER 2 ............................................................................................................................... 16

LITERATURE REVIEW ............................................................................................................... 16

2.1. DIGITAL FORENSICS ......................................................................................................... 16

2.1.1. Forensics - 'Defined' ...................................................................................................... 16

2.1.2. Digital Forensics Process ............................................................................................... 17

2.1.3. Digital Forensics tools .................................................................................................... 19

2.1.4. Digital Evidence, metadata and SOPs .......................................................................... 21

2.2. CLOUD COMPUTING ......................................................................................................... 24

2.2.1. Cloud ‘Defined’ ................................................................................................................ 24

2.2.2. Cloud deployment and service models ........................................................................... 26

2.3. DIGITAL FORENSICS FOR CLOUD ............................................................................... 30

2.3.1. Cloud forensics – ‘Defined’ ............................................................................................ 30

2.3.2. Cloud forensics – Processes ......................................................................................... 32

2.3.3. Digital forensics architecture and models for Cloud ....................................................... 34

2.4. SUMMARY .......................................................................................................................... 37

CHAPTER 3 ............................................................................................................................... 38

RESEARCH METHODOLOGY .................................................................................................... 38

3.1. RESEARCH PURPOSE ....................................................................................................... 38

3.2. RESEARCH APPROACH ...................................................................................................... 39

3.3. RESEARCH STRATEGY ....................................................................................................... 39

3.4. DATA COLLECTION ............................................................................................................ 41

3.5. DATA ANALYSIS ............................................................................................................... 42

3.6. RESEARCH QUALITY CRITERIA ........................................................................................ 42

3.7. SUMMARY .......................................................................................................................... 43
CHAPTER 4 ............................................................................................ .............................44
ANALYSIS AND FINDINGS........................................................................................................44

4.1. INTRODUCTION ..................................................................................................... .......44
4.2. LONG-TERM RETENTION OF DIGITAL DATA ..............................................................44
4.4. IMPLEMENTATION OF CLOUD-BASED DIGITAL ARCHIVING ..............................46
4.3. DIGITAL FORENSICS IN ARCHIVING ......................................................................51
4.5. ANALYSIS OF RESEARCH QUESTIONS .....................................................................53
   4.5.1. What are the different methods, tools and techniques in digital forensics? 53
   4.5.2. How can digital forensics be applied to capture digital evidence for ensuring authenticity, integrity and accessibility? ..........................56
   4.5.3. Which digital forensics methods and tools be used to capture digital evidence for digitized collections in Cloud environment? ........................59
4.6. SUMMARY .................................................................................................................66

CHAPTER 5 ............................................................................................ .............................67
CONCLUSION AND FUTURE WORK ........................................................................... ......67

5.1. CONCLUSION .............................................................................................................67
5.2. CONTRIBUTIONS TO THE RESEARCH .....................................................................68
5.3. SUGGESTIONS FOR FUTURE WORK .......................................................................68

BIBLIOGRAPHY .................................................................................................................70
List of Figures

Fig. 1 Outline of the thesis ..................................................................................................................15
Fig. 2 Digital evidence analysis as a process .......................................................................................21
Fig. 3 NIST Cloud Definition Framework ............................................................................................26
Fig. 4 NIST Cloud Computing Reference Architecture ........................................................................28
Fig. 5 Interactions between the actors in Cloud computing .................................................................28
Fig. 6 Service orchestration – Cloud provider ......................................................................................29
Fig. 7 Digital forensics process in Cloud environment ........................................................................33
Fig. 8 Forensics investigation process in Cloud environment ..............................................................35
Fig. 9 CDAC Cloud forensics model ...................................................................................................35
Fig. 10 Hyper-V Model ..........................................................................................................................36
Fig. 11 Step-by-Step Cloud-based Archiving Implementation ............................................................50
Fig. 12 Digital forensics in context of OAIS model .............................................................................52
Fig. 13 Forensics investigation process ...............................................................................................54
Fig. 14 Process-approach to digital evidence handling ......................................................................58
Fig. 15 High level Cloud forensics process .........................................................................................61
Fig. 16 Implementation of ISO standards in Cloud forensics .............................................................63
List of Tables

Tab. 1 Summary of Digital forensics process models..................................................18
Tab. 2 Sources and context of digital evidence ..........................................................22
Tab. 3 Situations of different research strategies.........................................................40
Tab. 4 Industry requirements for long-term retention of data ....................................45
Tab. 5 Cloud Service Providers for Digitization and Archiving.................................47
Tab. 6 Pros and Cons of Cloud for Digital Archiving................................................48
Tab. 7 Rules for digital forensics investigations .........................................................54
Tab. 8 Methods and tools in Digital forensics ............................................................55
Tab. 9 Top 10 Cloud threats and attacks .................................................................59
Tab. 10 Cloud forensics process phases and solutions ..............................................61
List of abbreviations

ACPO – Association of Chief Police Officers
ALM – Archives, libraries and museums
AWS – Amazon Web Services
API – Application programming interface
BCMS – Business Continuity Management System
BSI – British standards institution
CIO – Chief Information Officer
CISO – Chief Information Security Officer
CAGR – Compound annual growth rate
CSF – Critical success factors
CDAC – Center for development of advanced computing
CSP – Cloud service provider
DDoS – Distributed denial of service
DOJ – US Department of Justice
DNS – Domain Name Server
ECCC – European Convention on Cyber Crime
FTK – Forensic tool kit
FOSS – Free and open source software
FROST – Forensic OpenStack tools
HaaS – Hybrid as a service
HVAC – Heating, ventilation and air-conditioning
ISO – International standards organization
ICT – Information & Communication Technology
IoT – Internet of things
ISMS – Information Security Management System
IS – Information Security
MitM – Men in the middle attack
NIST – National Institute of Standards & Technology
OAIS – Open archive information system
PaaS – Platform as a service
SaaS – Software as a service
SOA – Service oriented architecture
SLA – Service level agreement
SOP – Standard operating procedures
SSL – Secure socket layer
VM – Virtual machine
VPN – Virtual private network
ZB – Zettabytes
Chapter 1
Introduction

This chapter covers introduction, background, research problem and the research questions to be answered in the study. The outline of the thesis and the brief coverage of the chapters is presented at the end of the chapter.

1.1. Background of the study

The amount of data has evolved multi-fold over the years both in terms of volume and variety. The advances in mobile computing, social networking, cloud computing and the storage technologies have further increased the flow of information and the accessibility across the organizations. With fast paced growth of digital technologies and ICT has already became an ingrained part of our everyday life, with everything from electricity to transportation and public safety being driven by the ICT solutions.

According to IDC (EMC Digital Universe Study, 2014), by 2020, the amount of information produced by machines, the so-called internet of things, will account for about 10% of data on earth. In 2013, only 22% of data was considered useful, even though less than 5% of that was actually analysed. By 2020, more than 35% of all data could be considered useful data, owing to the growth of data from the internet of things.

IDC has further predicted that the amount of data on the planet is set to grow 10-fold by 2020 to 44 ZB from around 4.4 ZB today. The growth in data is also coupled with the emergence of new business models for storing huge chunks of data in Cloud. The world is witnessing cloud shift (Gartner, 2016) in the previous 2-3 years which is predicted to grow enormously until 2020, complemented with a growth in Cloud data services, new IT architecture, storage models, operating philosophy, and new opportunities for digital business, and Internet of Things are fast emerging.

The maintenance of data or archive on cloud undoubtedly is a cost-effective solutions with complete host machines, data servers and operating systems hosted in Cloud. As the world is shifting to cloud, the pressure is also more on large organizations and public sector to adopt new technologies. More and more organizations are creating roadmaps that reflect the need to shift their IT strategy. It is predicted that by 2020, anything other than a cloud-only strategy for new IT initiatives will require justification at more than 30% of large-enterprise organizations (Panetta, 2017).
The fast adoption of cloud computing is definitely providing tons of benefits to the adopters with superior flexibility, accessibility, and capacity compared to traditional online computing and data storage methods (Lord, 2017). But, these benefits do come with essential requirements for ensuring better information governance, security policy and risk management to mitigate data security risks against cyber-attacks.

The mitigation of security risk is imperative to fulfill CIO and CISO expectations, to transition applications and data to the cloud platform. The applications, systems and data have different security threshold and the decision to migrate on the cloud platform is dependent on the sensitivity of data and the level of data security implemented. The CSF should be whether the value of data offsets the data security risk.

As the digital transformation initiatives picked the pace around the world, the threats and vulnerabilities for cloud environments have also grown enormously (Greef, 2017). The data stored in a cloud data storage such as Google Drive, Dropbox, Amazon Cloud Drive, etc. can be securely saved but the potential perpetrator may still hack into the system and delete or modify data without even being detected during the investigation. To mitigate such risks of data tampering, confidentiality breaches and cyber-attacks, the field of digital forensics provides several methods and tools to ensure capturing of digital evidence for the sustained integrity, authenticity and accessibility of data. The digital evidence capture can also lead to post-attack investigations and court trials.

The digital collections or e-records or digital archives also have specific objectives to ensure preservation of data objects, metadata and preservation description information (provenance, history etc.). The digital forensics tools can also help ensure preservation of essential data and aid in post-attack investigations, data recovery and litigation.

The research work in this thesis is focussed on investigating methods and tools in digital forensics and how they can be applied to the digital collections in cloud environment. The empirical investigation of Cloud platform (Google Drive) would help gather digital evidence to ensure authenticity, integrity and accessibility in the Cloud.

1.2. Problem statement

The amount of digital materials is increasing exponentially with the rapid development of digital epoch (Ma, 2008). But, all the digital information created every second worldwide is also getting lost, if not properly managed and secured. The new data storage models, file formats and operating platforms pose another threat to the
sustainability of digital data in new environments where the rendering software and the
data standards are rapidly changing posing another threat for obsolescence.

It was perhaps easier before to protect data in the data centers without essentially
requiring stringent security and forensics methods/tools to ensure data integrity. But,
now the new technologies are overflowing data outside via the cloud and mobile etc.
(EMC Digital Universe Study, 2014). For example, in 2018, 25% of company’s data is
predicted to come directly from IoT (from mobile to the cloud), and by circumventing
the security controls. Hence, the organizations today are now forced to respond to the
issues of security/risks from technologies and assets they no longer own or control.

Therefore, there is an apparent need to have an efficient digital forensics methods and
tools understanding amongst custodians of digital data in the memory institutions and
large organizations to securely maintain digital repositories. They must understand the
complete process as to how the evidences are gathered and as how the authenticity,
integrity and accessibility of digital data in physical or cloud storage be maintained.

There have been a lot of research done in cyber security, digital forensics and the
preservation of digital data. But, there has not been a lot of research found focussing
on digital forensics in context of electronic records or digital documents/record in cloud
environment. One research project which was found closest to the current research
project is BitCurator Access project of University of North Carolina (BitCurator, 2016),
which focused on approaches to simplify access to raw and forensically-packaged disk
images; allowing collecting institutions to provide access environments that reflect as
closely as possible the original order and environmental context of these materials.
The use of forensic technologies allows for detailed metadata to be generated
reflecting the provenance of the materials (Lee, 2012), the exact nature of the file-level
items they contain, and the metadata associated with both file-level items and data not
observed within the file system (but still accessible within the original materials).

1.3. Purpose of the study

The reliability of digital data is of paramount importance for the organizations especially
when the data is either mission critical or of high cultural significance. With a
transforming cloud based data storage models (HaaS, SaaS, PaaS etc.), the
organizations are rapidly migrating their digital data to the new platforms. But, despite
of ease of use and cost advantage, the most important concerns for the organizations
hosting their data on cloud are data security and business continuity. Hence, there is
a demand for developing proper governance, security policies and digital forensics to ensure security of digital data and gathering of evidences for integrity and accessibility.

The research objectives of this thesis is to conduct research on digital forensics and how it can be handy in ensuring security and gathering evidences for authenticity, integrity and accessibility of digital data hosted in the cloud-based storage.

The empirical investigations will be conducted using state of the art forensics tools to demonstrate the applicability of digital forensics (also cloud forensics) on digital collections hosted in cloud storage. The research presented in the thesis work would be useful for the organizations willing to implement digital forensics tools and methods.

The research objectives are also to promote the discussion on digital forensics amongst the custodians of digital data (digital manager, document manager, records manager, librarian, archivists, knowledge managers etc.). It may help them take decisions on adopting cloud storage for hosting document repositories, data archive, digital libraries and knowledge management portals.

1.4. Research question

In order to achieve the objectives and the purpose of the study, the following major research questions are posed:

1. What are the state-of-the-art methods, tools and techniques in digital forensics?
2. How can digital forensics be applied to capture digital evidence for ensuring authenticity, integrity and accessibility?
3. Which digital forensics methods and tools be used to capture digital evidence for digitized collections in Cloud environment?

1.5. Research strategy

The field of digital forensics is focussing mainly on the gathering of empirical evidences for the reported incidents or crimes. There is a detailed conceptual view in digital forensics for problem investigation and there are plethora of tools to ease the process of deducing conclusions. In this thesis, it is observed and highlighted the increased migration of digital documents of the institutions to the highly economical and scalable cloud environments. There are many such platforms available (such as Google drive, Dropbox etc.). However, the transition to Cloud has been very slow in the memory institutions (ALM), therefore there has not been any site identified for the case study.
The empirical study would be conducted on the defined sample digital collection. The qualitative methods were applied to conduct the study. The empirical evidence were gathered from the tests conducted on the pilot system (Google Drive). Finally, the research methodology and the discussions would be presented in the next chapters.

1.6. Relevance of study

The organizations are increasingly migrating their data, applications and infrastructure to Cloud. Due to cost, scalability and ease of handling, the organizations are rapidly transitioning to virtualization. The mass-benefits of cloud and virtualization comes with the trade-offs for security, privacy and data integrity. Therefore, for ensuring the data security and continuity in cloud-based storage, the organizations requires necessary methods/tools for digital forensics to ensure sustained authenticity, integrity and accessibility. This thesis work is relevant to such type of work because it can help organizations define required strategies to gather digital evidences for the sustained storage/preservation of digital data in the cloud environment.

1.7. Limitations and delimitations

The research on digital forensics applications to cloud environment “cloud forensics” is quite new. There has not been any research identified focussing on the applications of “cloud forensics” to the digital collections hosted in cloud environment.

The span of research is also quite wide which has to be covered in the limited amount of time. Therefore, it was chosen to focus only on the sample pilot collections for empirical investigations instead of conducting a case study in the organizations.

The application of digital forensics to digital preservation and archiving is being researched in the BitCurator project. The research is still undergoing and there has not been a lot of information available on the findings and the dedicated tools developed for the memory institutions and ALM. Hence, only the openly available forensics tools were used to conduct the investigations and derive conclusions for digital collections.

1.8. Thesis structure

The thesis is described in seven chapters.

Chapter 1 introduces the study with brief introduction, background, problem statement, purpose, research questions, limitations and delimitations of the study.

Chapter 2 represents literature review and the related works relevant to thesis work.
Chapter 3 provides the methodology used to obtain data, its analysis, the factors taken in to consideration, and the problems faced during the research.

Chapter 4 deals with the findings and interpretation of the results with respect to the research questions.

Chapter 5 contains the conclusion and recommendations for future research.

Fig. 1 Outline of the thesis
Chapter 2.
Literature Review

This chapter covers the theoretical background of digital forensics, cloud computing and the cloud forensics applications to digitized collections in cloud. The research work builds upon the theoretical foundation to conduct detailed analysis and the empirical investigations on the topic.

2.1. Digital forensics

2.1.1. Forensics - ‘Defined’
There have always been crimes leading to the investigations, nabbing of culprits and litigation. Over the years there have been growth in crimes as well as the growth in the science of “forensics” which deals with the investigations and the court proceedings. The science of “forensics” is not new but with an advent of digital era, it has evolved to a new level with “cyber-piracy”, “hacking”, “online frauds” and “cyber-attacks”.

The digital devices are increasingly being used to commit crimes or used as an accessory to commit crimes. It may be quite easy to gather evidence from a physical crime scene but in digital realm it proves to be more and more difficult. The dynamic nature of technology, complexity and increased number of security breaches requires sophisticated forensics. That's why digital forensics has become increasingly popular over the last decade due to the increased presence of digital evidence in courts across jurisdictions in both criminal and civil cases (Cohen, 2009).

Digital Forensic Research Workshop (Palmer, 2001) defined digital forensics as:

“The scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or gathering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations”.

Another definition of digital forensics defines it as (Willassen, 2005):

“The practice of scientifically derived and proven technical methods and tools towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of after-the-fact digital
information derived from digital sources for the purpose of facilitating or furthering the reconstruction of the events as forensic evidence”.

The common elements from the definitions and the discussion has identified common elements which may be elaborated as follows:

- Applied to digital media and/or digital data
- Scientific and proven methods and tools
- Pre-defined and/or accepted process
- Consider legal principles
- Extract digital evidences
- Indicate a set of events / actions being the root cause.

Digital forensics is a specific, predefined and accepted process applied to digitally stored data or digital media that use scientific proven and derived methods, based on a solid legal foundation, to produce after-the-fact digital evidence (Palmer, 2001).

### 2.1.2. Digital forensics process

The digital forensics investigations have to follow a standardized approach to conduct investigations and to gather digital evidences. However, to data, there has not been any single, standardized and consistent digital forensics investigation process model which has been unilaterally accepted worldwide (Ngobeni, 2016).

Numerous scholars and researchers have created various digital forensic investigation process models. This section compares and contrasts various digital forensics process models. Amongst various models, DFRW is one of the participants that took the initiative to develop a consistent and standardised digital forensic process model (DFRW, 2001). The greatest challenge in respect of this process model is that the analytical procedures and protocols are not standardised, nor do practitioners and researchers use standard terminology (Reith, Carr, & Gunsch, 2002).

However, the general terminology to define digital forensics has been widely agreed as a specific, predefined and accepted process applied to digitally stored data or digital media that use scientific proven and derived methods, based on a solid legal foundation, to produce after-the-fact digital evidence (Palmer, 2001).

A specific physical crime scene investigation procedure was proposed to investigate crimes (James S. H & J., 2005). Their model highlights the importance of evidence that can be gathered at the crime scene. The model show through its documentation phase
that the physical crime scene should be properly documented and the digital forensics investigators should use their expertise to gather useful pieces of evidences. Though it was a good starting point for a viable digital forensics model, it is not only the physical objects that can be found at the crime scene.

**Tab. 1 Summary of Digital forensics process models**

<table>
<thead>
<tr>
<th>BSI/ISO 27043</th>
<th>Kent et al.</th>
<th>James and Nordby</th>
<th>US DOJ</th>
<th>DFRW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.Crime scene reconstruction</td>
<td>5.Investigation of the incident (data collection and analysis)</td>
<td>5.Analysis</td>
<td></td>
</tr>
</tbody>
</table>

A four stage digital forensics process model was proposed by NIST with collection, examination, analysis, and reporting as the four stages (Kent, 2006). It was one of the most basic and simple digital forensics model to follow by the organizations and it was well elaborated in the NIST standard. However, it also possesses the same shortcomings as that of the process model developed by Reith and US DOJ which uses examination and analysis to identify and collect digital evidences.

The BSI/ISO 27043 proposes a harmonised digital forensic process model which includes four stages: readiness, initialisation, acquisitive and investigative (ISO, 2015).
This ISO standard describes the overall process to be followed when conducting digital forensic investigation. It is a good standard but the major challenge is that it is not encompassing, it does not cover all cases of digital forensic investigations from a general hard drive acquisition to a more advanced network intrusion forensics.

Despite of the several digital forensics process models existing (as highlighted in the Table 2.1), there has not been any consensus about a single, standardized digital forensics process model that can be predominantly adopted globally.

2.1.3. Digital forensics tools

The digital forensics is the process of recovering and preserving materials found on digital media / devices. It is needed because data are often deleted, locked or hidden. The digital forensics tools are hardware and software tools that can be used to help in the recovery and/or preservation of digital evidence.

The tools help practitioners to capture reliable digital evidences which could be presented in the court. The digital forensics tools provide the investigator with access to the evidence but typically do not provide access to methods for verifying that the evidence is reliable (Carrier & Spafford, 2006).

The law enforcement uses both digital forensics software and hardware, mostly open source or commercial tools are used, depending on the type of cyber-crime.

**Hardware:** The hardware tools are primarily used for storage device investigations, to keep the devices unaltered in order to preserve the integrity of digital evidence. For example, a hard-drive duplicator is an imaging device that copies all files from a suspected hard drive onto a clean drive; A password recovery device employs algorithms such as brute-force or dictionary attacks to attempt password cracking for the protected storage devices; A forensic disk controller is a read only device that allows the user to read data from the suspected device without altering data integrity.

**Software:** Many of the forensics applications are included with BackTrack or Kali Linux distributions. There are many applications available, open-source and commercial, allowing digital forensics practitioners to performance digital evidence gathering and data recovery. The law enforcement complements the hardware tools with advanced software solutions, a hardware tool such as write-blocker could preserve evidence in the target device whereas the software tool can acquire and analyse digital evidence collected from the target device. The cyber criminals often hide files, partitions or hard drives to masquerade gathering of evidences; however, the forensics software can
assist in recovering the hidden or deleted files. The windows registry records creation, modification and deletion of files, there are forensics software available which can perform registry analysis and collect traces of activities performed. In summary, system and user activities can be recovered and investigated using forensics software.

The digital forensics tools are classified based on their role in the digital forensics process. The tools are often developed for a specific device or for a specific operating systems. The role of the tools include: evidence acquisition, evidence examination, evidence analysis and integrated tools (Hewling, 2013). The details of groups include:

1. **Acquisition tools:** The digital forensics acquisition tools are the set of tools that are used to create a mirror copy or image of the target device. The cryptographic hash is usually made at the time of acquisition which is one of the key action for maintaining the chain of custody of the evidence. This phase of digital investigation process is very important and the key objective here is to preserve the integrity of the target device. It helps in maintaining the integrity of the digital evidence and ensure chain of custody. The digital forensics acquisition tools are often used in conjunction with write blockers to ensure that nothing is written to the target device during the digital forensics investigation process. But, despite of all good intents, how does the practitioner may ensure that the target device is not tampered or intentionally or accidently during gathering of evidence. The answer to this lies in the integrity tools which are being used by the practitioners to ensure data integrity. This research work investigates the techniques / tools used to ensure preservation of integrity for the target device.

2. **Evidence examination and analysis tools:**
   The tools for examination and analysis are used to extract and analyse the digital evidence. The extraction are of two types: 1) physical extraction recovers all the data from the drive regardless of the file system, and 2) logical recovers files based on the devices, operating system and file systems (Goodison, 2015).

3. **Multipurpose (integrated) tools:**
   These are the tools with different functionalities integrated into one tool. These tools carry out multiple processes from search, data acquisition, navigation, extraction, examination, analysis and reporting. These tools are commercial (e.g. Encase or FTK) developed for specific devices or interfaces. The data acquisition feature of these tools facilitates copying or producing mirror image of the system or the target device to be investigated. The search feature of these
tools facilitates identification of data that matches particular criteria, ranges or classifications pre-selected by the digital forensics expert. The navigation feature of the tool facilitates exploring of a digital crime scene and help digital forensics expert to visualize the crime scene (Schatz, 2007).

The extraction feature of tool facilitates extraction of data from applications running on the target device such as internet browser artefacts. The examination and analysis tools facilitates in gathering valuable relevant information for the captured digital data. The integrated tools (Case Management tools) have emerged as a one fit for all solution to the growing problem of increasing data volume that contains potential evidences. Integrated tools provides various ways of searching, filtering and analysing of digital data gathered from the crime scene. The tools such as FTK and Encase provides various ways to assist digital forensics experts to produce verifiable digital evidence.

There are number of digital forensics tools available, both free and commercial, developed by various organizations and groups with different purposes. Despite of the fact that these tools provides underlying basis to the field of digital forensics, the field is open to disparities. This research study aims to present the essential pragmatic tools and methods of digital forensics useful for memory institutions and ALM sector.

2.1.4. Digital evidence, metadata and SOPs
The basic entity of any cyber related crime lies in the “digital evidence”. This evidence is captured from a crime scene through the digital forensics process, which is an investigative process comprising of collection, preservation, interpretation and presentation of evidence. The process followed by one investigator may differ from another despite of various legislations in place. But, the corresponding digital evidence may be unique based on the form it take which is not necessarily the physical one.

![Fig. 2 Digital evidence analysis as a process](image)
The term digital evidence is defined as, “Encompasses any and all digital data that can establish that a crime has been committed or can provide a link between a crime and its victim or a crime and its perpetrator” (Casey, 2011). The term may also be defined as, any data or information found to have been stored or transmitted in a digital form that may be used in court (Hewling, 2013). Digital evidence may also be referred as a bag of bits, which in turn can be organized as the sequences to represent information. The information in sequential bits will seldom make sense and tools are required to represent these structures logically in a human readable form (Cohen, 2009). The process view of digital evidence (refer to Fig. 2) includes discovering, capturing and processing of digital evidence in order to be presentable in court (Rahurkar, 2012).

There are various tools and methodologies used by digital forensics experts but there is not any one internationally accepted benchmark. “This lack of formalization makes it more difficult for courts and other decision makers to assess the reliability of digital evidence and the strength of the digital investigators’ conclusions” (Casey, 2011). Furthermore, it is stated that that: “The number one problem in current litigation is the preservation and production of digital evidence” (Fulbright & Jowoski. L., 2006).

The preservation is characterized as one of the key process in the digital forensics.

**Tab. 2 Sources and context of digital evidence**

<table>
<thead>
<tr>
<th>User-created files</th>
<th>User-protected files</th>
<th>Computer-created files</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Address book</td>
<td>• Compressed files</td>
<td>• Backup files</td>
</tr>
<tr>
<td>• Database files</td>
<td>• Misnamed files</td>
<td>• Log files</td>
</tr>
<tr>
<td>• Media files (images, audio, video etc.)</td>
<td>• Encrypted files</td>
<td>• Configuration files</td>
</tr>
<tr>
<td>• Text files (.doc, .xls, .ppt)</td>
<td>• Password protected files</td>
<td>• Printer spool files</td>
</tr>
<tr>
<td>• Bookmarks / favorites</td>
<td>• Hidden files</td>
<td>• Cookies</td>
</tr>
<tr>
<td></td>
<td>• Steganography</td>
<td>• Swap files</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• System files</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• History files</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Temporary files</td>
</tr>
</tbody>
</table>

The sources and context of capturing digital evidence (Tab. 2) aids the investigation process (EC Council, 2017). The metadata is used to capture details about the digital evidence and it is helpful in determining the context of the digital evidence. Metadata increases the information about the content and may even extend the context in which the data was captured. Metadata aids the digital forensics expert during an investigation to derive number of inferences about the digital data captured from the
target device. The metadata may give indications about the following key parameters (Köhn, 2012):

1. Data (file/folder) creation date, last accessed, last modified
2. File/Folder storage path
3. File/Folder function (examined by the extension and header data)
4. File/Folder owner and inherited permissions
5. File/Folder size
6. File/Folder hash signature

The digital forensics expert often have to explain as to how the digital evidence was collected. The national security organisations and governmental authorities have formulated standard operating procedures and best practice guidelines dealing with the discovery and capture of digital evidence. There are various tools and technologies available for the investigators to successfully capture and present digital evidence to the law enforcement authorities. The tools and technologies have been developed based on the requirements gathered from the cyber-crime incidents. The methods are refined during the process and subsequently used for other investigations. One of the major requirements for digital evidence is authenticity (i.e. data must be from authentic source). Digital evidence is verified by hashing the data content (Carrier B., 2002).

The method or process is always emphasised as critical during a digital forensics investigation. It is stated that every piece of digital evidence should be challenged to ensure that an investigator followed a rigorous process (Ruibin & Gaertner, 2005). The SOPs form part of the digital forensics process which are continually been evolving with emerging technologies. The industry specific or general SOPs are often derived from the best practice principles. Two of the best practice principles which are most frequently been refereed in digital forensics are:

1. Good Practice Guide for Computer Based Electronic Evidence or ACPO guide developed by the Association of Chief Police Officers in the United Kingdom
2. European Convention on Cyber Crime or ECCC guide

The best practice principles listed in the ECCC guide are similar to the ACPO guide (DOJ, 2001). The 4 best practice principles listed in the ACPO guide are (7Safe, 2008):

1. The data on the device is not to be changed by any action during investigation.
2. In specific circumstances, it may be necessary to access the original data on the device, in such cases a competent person with necessary expertise must do so and will be liable to explain the relevance and implications of the actions.

3. A clear audit trail, log, chain of custody or other records of all the processes applied to the digital evidence should be created and preserved. The processes should be reproducible by an independent third party to get the same results.

4. The investigating officer in-charge of the investigation is responsible to ensure that the aforementioned law and the principles are adhered to.

The best practice guidelines, SOPs and specific digital forensics processes help ensure that the digital evidence is preserved for authenticity, integrity and accessibility at all times during the investigation. The specific digital forensics software and the methodologies are utilized to ensure that the digital evidence is captured, preserved and presentable in the court without being tampered, damaged, or degraded. This research work investigates the applications of specific digital forensics software and methodologies (such as processes, metadata, SOPs etc.) to ensure the authenticity, integrity and accessibility of digitized collections / electronic records in Cloud platforms.

2.2. Cloud Computing

2.2.1. Cloud ‘Defined’

Cloud computing is a new field in the rapidly growing computing industry. It is continuously evolving field and the corresponding definitions are also expanding in scope and scale. One of the most well-known definition of cloud computing was given by US National Institute of Standards and Technology (Mell & Grance, 2011):

“Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models”.

The other definition of Cloud computing defines it as: “A style of computing where scalable and elastic IT capabilities are provided as a service to multiple customers using internet technologies” The definition from Gartner covers parts of the characteristics of Cloud computing but it does not contain references to on-demand
services as well as any pay-as-you go usage model. This implies that the definition does not consider these characteristics fundamental to the Cloud computing model.

The definitions mentioned above cover many technologies and various models.

The five essential characteristics of cloud model explained by NIST are:

1. **On-demand self-service**: A user can get need based automatic provisioning of computing capabilities without requiring human intervention.

2. **Network access**, by which services are available through the network promoting usage by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

3. **Resource pooling**, which enables pooling of computing resources to serve multiple clients using a muti-tenant model, with multiple physical and virtual resources dynamically assigned and reassigned based on demand. Examples of resources include storage, processing, memory, and network bandwidth.

4. **Rapid elasticity**, by which the Cloud services can provision resources rapidly and can scale up or down the resources based on client’s demand.

5. **Measured service**, enables monitoring, controlling, optimizing and reporting of services (or resources) based on metering capability for botch Cloud provider and the customers in order to ensure transparency.

Cloud computing has captured the mainstream over the past 3-4 years. The major IT companies such as HP, IBM, Intel, Microsoft and SAP have evolved their business strategies around cloud computing (IMDA, 2017). The key software houses such as Microsoft and Oracle are offering their software suites around cloud services to address the growing demand for utility based charging and collaboration.

The emergence of Cloud-aware application design patterns is making the development of Cloud based applications more convenient. The concept is now to focus on the idea rather than focussing on programming. Furthermore, the improvement of the instrumentation offered by the standardized interfaces of both Cloud infrastructure and Cloud platforms make application development and deployment more convenient (Khan, 2014). Cloud computing is in essence an economic model—a different way to acquire and manage IT resources. Organizations typically adopt cloud computing as a way to solve a business problem and not a technical problem (Lewis, 2017).

Cloud computing evolved from a myriad of technologies such as autonomic computing, grid computing, multi-tenancy, service oriented architecture (SOA) and virtualisation.
It outlines the underlying architecture upon which the services are designed and applied to utility computing and data centres. It abstracts the design details from the cloud user in order to present computing as an on-demand service.

The Cloud services and their delivery are the core of Cloud computing. The primary focus of the Cloud computing model is on economic method of providing higher quality and faster services at a lower cost to the users. Whereas the traditional service delivery model focussed mainly on procuring, maintaining and operating the necessary hardware and related infrastructure. The Cloud computing model enables the enterprises to direct their attention to innovative service creation for the customers.

2.2.2. Cloud deployment and service models

The NIST definition of Cloud computing (refer to Fig. 3) is a widely accepted and a valuable contribution towards providing a clear understanding of cloud computing technologies and cloud services (Mell & Grance, 2011). It provides a simple and unambiguous taxonomy of three service models available to cloud consumers: cloud software as a service (SaaS), cloud platform as a service (PaaS), and cloud infrastructure as a service (IaaS). It also summarizes four deployment models describing how the computing infrastructure that delivers these services: private cloud, community cloud, public cloud, and hybrid cloud (Lie & Jing, 2011).

![Fig. 3 NIST Cloud Definition Framework](Source: NIST)
The NIST cloud computing reference architecture focuses on the requirements of “what” cloud services provide, not a “how to” design solution and implementation. The reference architecture is intended to facilitate the understanding of the operational intricacies in cloud computing. It does not represent the system architecture of a specific cloud computing system; instead it is a tool for describing, discussing, and developing a system-specific architecture using a common framework of reference (Lie & Jing, 2011).

Cloud deployment affects the scale and efficiency of the cloud implementation. Hence, one or more of the Cloud deployment models are implemented by the organizations depending on their requirements in terms of application, infrastructure, network and security. The Cloud deployment models followed in the industry are:

1. **Private Cloud**: It is a cloud infrastructure operated solely for a single organization. These single tenant clouds may be managed by the organization themselves or a third party and may be hosted within the organization or in a third party data center.

2. **Public Cloud**: It is a cloud infrastructure operated by a cloud provider which is available for the public hosting. These multi-tenant clouds serve a variety of customers and usually have largest scale and utilization efficiency. AWS and Microsoft Azure are two well-known public cloud providers.

3. **Community Cloud**: It is a public cloud infrastructure catering to the industry or community specific requirements (e.g. security and compliance requirements or common application requirements). SITA ATI Cloud provide services to airline employees for infrastructure, desktop and other services.

4. **Hybrid Cloud**: It is a cloud infrastructure deployed across two or more cloud deployment models. A successful hybrid cloud implementation model requires integration that enables data and application portability between different cloud services. The most common hybrid clouds comprises of private and public clouds where the workload is over-flowed from private into public cloud.

The NIST cloud computing reference architecture defined five actors: cloud consumer, cloud provider, cloud carrier, cloud auditor and cloud broker (Lie & Jing, 2011). Each actor of these actors is an entity (a person or an organization) participating in a transaction or process and/or performs required tasks in Cloud computing.
The interaction among the actors in the Cloud setup provides advantage of significant economies of scale in three areas:

1. Supply-side savings in cost per server;
2. Demand-side aggregation increases utilisation; and
3. Multi-tenancy efficiency distributes application management and server costs

For example, consider the Microsoft’s 700,000 square-foot Chicago Cloud Data Centre. The Microsoft facility currently houses 224,000 servers in 112 forty-foot containers, and has a maximum capacity for 300,000 servers. This US$500 million facility is operated by a skeleton crew of only 45 (IMDA, 2017).
The communication path in Fig. 5 illustrates the interaction between different actors in the NIST Cloud computing reference model. A cloud consumer may request cloud services from a cloud provider directly or via a cloud broker. A cloud auditor conducts independent audits and may contact the others to collect necessary information. The complete interaction may be understood by the following scenarios (Lie & Jing, 2011):

**Scenario 1:** The cloud consumer may request service from a cloud broker instead of contacting cloud provider directly. The cloud broker may create a new service by combining multiple services or by enhancing existing service. The cloud providers are invisible in this example and the cloud consumer interacts directly with cloud broker.

**Scenario 2:** The cloud consumer gets connectivity and transport of cloud services from cloud carriers. Cloud provider may sign two SLAs, one with a cloud consumer (SLA1 and another with cloud carrier (SLA2). SLA1 may detail out the essential requirements and SLA2 may specify the requirements for capacity, flexibility and functionality.

**Scenario 3:** The cloud auditor conducts independent assessments of the operation and security of the cloud service implementation. The audit may involve interactions with both the cloud consumer and the cloud provider.

![Fig. 6 Service orchestration – Cloud provider](image)
As mentioned in the NIST cloud computing definition (Mell & Grance, 2011), a cloud infrastructure may be operated in one of the following deployment models: public, private, community and hybrid. The differences are based on the deployment on resources accessible to the Cloud consumer. A three-layered model is used to understand the service orchestration, representing the grouping of three types of components Cloud providers need to deliver their services (Lie & Jing, 2011):

The top layer in the model is service layer, this is where Cloud providers define interfaces for Cloud consumers to access the services. The access interfaces of each of the three service models (PaaS, SaaS, and IaaS) are provided in this layer. The middle layer in this model is the resource abstraction and control layer. It contains the system components that Cloud providers use to provide and manage access to the physical computing resources. The examples of resource abstraction components include hypervisors, virtual machines, virtual data storage and other abstractions. The control aspect of this layer include software components that are responsible for resource allocation, access control and usage monitoring. The lowest layer in the stack is the physical layer which includes all the physical computing resources. It includes hardware resources such as computers, networks, storage and other physical devices. It also includes facility resources such as HVAC, power, communications etc.

The different cloud deployment models were highlighted in Fig. 6. But, the models do have important security implications. For instance, a private cloud is dedicated to one cloud consumer, whereas a public cloud could have unpredictable tenants. Therefore, workload isolation is less of a security concern in a private loud than in a public cloud.

This thesis work investigates the process approach of applying digital forensics to the digital collections or organizational electronic records hosted in Cloud platform. These collections may be hosted using one of the deployment models utilizing one or more of the service models. Since, the security is one of the major concerns in any of the cloud deployment, hence, it becomes necessary to explore the necessary digital forensics tools and methods for ensuring authenticity, integrity and accessibility.

2.3. Digital Forensics for Cloud
2.3.1. Cloud forensics – ‘Defined’

Over the years since 2010, the cloud computing has revolutionized the methods and the tools through which digital data is stored and transmitted. On one side, it has made easy for the companies to manage their storage and infrastructure, whereas on the
other side, it has aggravated many technological, organizational and legal challenges. Several of these challenges, such as those associated with data replication, location, transparency, and multi-tenancy are somewhat unique to cloud forensics (NIST, 2014).

**NIST defines cloud computing** (Mell & Grance, 2011) as “a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources (e.g., network, server, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model composed of 5 essential characteristics, 3 service models, and 4 deployment models.” **Cloud forensics is a process applied to an implementation of this model.**

A pragmatic definition of cloud forensics defines it as the application of digital forensics science in the cloud environments (Ruan, Carthy, & Crosbie, 2011). It consists of a hybrid approach (e.g. remote, virtual network, large-scale, thin client, and thick client) towards generating digital evidence. Organizationaly, it involves interactions among cloud actors (cloud consumer, cloud provider, cloud broker, cloud carrier, cloud auditor) for facilitating both internal and external investigations. The investigations often implies multi-jurisdictional and multi-tenant situations.

A working definition of cloud forensic defines it as “the application of computer forensic principles and procedures in a cloud computing environment” (Shams & Hasan, 2013). The two main parties involved in cloud forensics investigations are cloud consumer and the cloud provider. Many cloud providers outsource their services to third parties, thus increasing the scope of the investigation, in the case of a crime. Today, the definition of cyber-crime have extended to include cloud crime as (Saha, 2017),

1. **Cloud as an object:** When the target of the crime is either the cloud provider or the cloud broker and the cloud is attacked in parts or as a whole.
2. **Cloud as a tool:** When the data related to the crime is hosted or saved on the cloud server and the cloud network is used to facilitate cyber-crimes.
3. **Cloud as a subject:** When the crime is committed within the established cloud environment.

Cyber criminals may use DDoS attack to target cloud provider, or use cloud environment to commit crime such as identity theft of the cloud user, illegal access and/or tampering of data hosted in the cloud, or use cloud as a platform to store crime related data and share among culprits (Dobrosavljević & Veinović, 2015). One of the
major challenges in the cloud forensics is how to identify the digital evidence. For example, in IaaS, it is easy to determine the data location on the servers if the data is located on the direct attached storage. However, with a progress in virtualization, more and more servers do not have a direct attached storage but mapped storage devices and it has increased complexity specifically in context of virtualized cloud environment.

Cloud forensics is still in its infancy; despite dozens of articles in the literature over the last five years, there is a notable dearth of usable technical solutions on the analysis of cloud evidence (Roussev, Ahmed, Barreto, & McCulley, 2016). Moreover, we are still in a phase where the majority of efforts are targeted at enumerating the problems that cloud poses to traditional forensic approaches, and investigating for the ways to adapt (with minimal effort) existing techniques. This thesis work tries to investigate the domain of cloud forensics and understand as to how it applies to the digital collections or e-records or digital archives hosted in the cloud-based storage.

2.3.2. Cloud forensics – Processes
The application of digital forensics has not been so easy task in context of cloud computing. According to Gartner, “Cloud services are especially difficult to investigate, because data access and data from multiple users can be located in several places, spread across a number of servers that change all time” (Gartner, 2010).

The cloud computing technology is a dynamic service-oriented approach. It creates many challenges to the applicability of existing digital forensics procedures to cloud environment. Hence, the new process model was prepared by the researchers encapsulating DFRWS and NIST process models as well as the IDIP. The process model for cloud forensics includes following stages (Almulla & Iraqi, 2014):

1. **Identification**: The sub-process of identification determines the type of crime, software and hardware used by the suspect and the possible digital evidence locations. In cloud environment, the identification of digital forensics requirements to conduct a sound investigation is considered to be the main building block of this sub-process.

2. **Preservation**: The sub-process of preservation ensures digital evidence integrity by preserving the integrity of the original data. In cloud environment, the challenge is how to preserve the data and then determining whether the existing approaches of data integrity (e.g. hash function) are applicable or not.
3. **Collection**: The sub-process of collection extracts the exact bit-by-bit image of the required data. In cloud environment, the collection of whole target environment might not be possible due to the fact that the infrastructure is outsourced and owned by the cloud provider. Furthermore, the variations of cloud service models pose new set of challenges on evidence collection.

4. **Examination**: The sub-process of examination study the collected data and attributes. The current digital forensics practices emphasizes on examining the well-structured storage e.g. hard disks. In cloud environment, it is complex to examine the storage because the significant proportion of target data is held in memory / network dumps and / or log files.

5. **Analysis**: The sub-process of analysis conducts in-depth systematic evidence search on target devices: live and/or static systems. In cloud environment, the forensics expert must consider the dependencies of a cloud based application on the service provided within cloud provider boundaries or outside. If the complete chain of custody is not possible, then investigators need to perform analysis on the partial sources in hand (Almulla, Iraqi, & Jones, 2011). Further, there is also a need for a digital forensics tool capable of acquiring and analysing
cloud-based cases e.g. FROST (Dykstra & Sherman, 2013). Many tools are available to perform analysis such as, EnCase and Forensic Tool Kit (FTK).

6. **Presentation:** The sub-process of presentation prepares summarized findings or report to present it to either management or organization or a court of law.

The process approach simplifies the complexity of investigations in the cloud environment. However, the rapid growth of cloud technology does require frequent adaptations in the process model for cloud forensics to include new changes. It is mandatory for the digital forensics expert to carefully conduct the identification phase to determine if it is really a cloud environment or any other form of web service, or even a VPN. An error in the identification phase may lead to investigation failure or violation of standards and best practices. This research work investigates the process model and the existing architecture or framework of digital forensics for its applications to the digital collections or e-records hosted in the cloud environment.

### 2.3.3. Digital forensics architecture and models for Cloud

The most challenging part is to incorporate the traditional models of digital forensics into black box architecture of cloud computing. There have been a lot of research undergoing on how to apply new methods, models and tools of security / forensics on cloud platform. The researchers are developing new models, methods and architectural frameworks to conduct digital forensics for cloud environments. One of the generic action oriented process model for applying digital forensics to cloud environment is illustrated in the following Fig. 8.

The new process model for forensics as mentioned in Fig. 8 differs significantly from that of the traditional forensic process. Firstly, the process step identify the purpose of the investigation followed by the type of cloud services (SaaS, PaaS, IaaS), and then the type of device, software and platform. The technology behind the concerned cloud is always verified so that the specific investigation process can be executed smoothly.
In view of the numerous challenges and multitude of attacks faced by the cloud platforms, a framework was proposed to implement forensics in the cloud environment. The framework was developed by CDAC to implement forensics in cloud platform (Datta, Majumdar, & De, 2016). The complete architecture has been divided into 4 layers: abstract layer, front-end, middle-end, and back-end.
The front-end is mainly an interacting layer of the model with a main component an API interface. Middle end is concerned with the database and maintains all the relevant data for forensics process. Backend component has several data mining techniques to segregate the relevant evidences which are the proofs for a specific crime scene and delivers those digital evidences to the presentation layer.

![Fig. 10 Hyper-V Model](source)

A model incorporating the findings of many of the existing reference models have been developed, named as Hyper-Model (Marangos & Panagiotis, 2012). The model is holistic in approach and is capable of mapping all the processes of cloud forensics technique. The model is modular and comprises of three basic modules: preparation, investigation, and presentation. Each of these phases are further sub-divided into sub-modules as illustrated in Fig. 10. The main objective of several phases is to identify the proper host, and then collect data from the particular host.

All the relevant digital evidences are then presented by preserving all their integrity and authenticity without being modified or tampered. The preparation phase comprises of identification, preparation and approach strategy whereas the investigation phase comprises of preservation, collection, examination and analysis. Finally, the presentation phase deals with provenance of the evidences from identification to the court of law. The preservation and provenance (chain of command) are the most important steps of the process from the authenticity, integrity and accessibility point of
view. The model have provided a process-view as well as a broader overview of the cloud forensics. It has provided an important background to investigate the topic of digital forensics for cloud. The research work presented in this thesis is built upon the background and the theoretical foundation presented in this chapter. Next chapters will present methodology, analysis, findings, discussion and final conclusion of the thesis.

2.4. Summary
The literature review presented in this chapter has covered the background of digital forensics, cloud computing and the applications of digital forensics to cloud environment. A brief introduction to the forensics tools, digital evidence, metadata and the SOPs was defined. The introduction to cloud computing with the deployment and the service models was briefly explained. The chapter ended with the overview of cloud forensics processes, architecture, framework and models. The chapter prepared the necessary background work for conducting the research work for the Master thesis.
Chapter 3
Research Methodology

This chapter covers research strategy and the methodology chosen for conducting research work. The research work builds upon the theoretical foundation developed through literature review. The method for collecting data and empirical investigations is based upon research problem/questions.

3.1. Research Purpose

The selection of research instruments and the adoption of appropriate methods paves the pathway to construct the research setting. It is the selected research instrument that suggests, how to solve a problem and reach new knowledge. Each research project has its own purpose and it is always possible to classify and base the research on that purpose. A research setting may be of either exploratory, descriptive or explanatory or a combination of these (Saunders & Lewis, 2012).

Exploratory research: It has the goal of formulating the research problem more precisely, clarifying concepts, gathering explanations, and forming hypothesis. This form of research can be performed using literature research (secondary research), survey, focus groups and case studies. The exploratory research may form the basis to develop the hypothesis, but it does not seek to test it (Darabi, 2007).

Descriptive research: It has the goal of establishing the answers to who, what, when, where and how? The accuracy is the major part of the importance in case of descriptive research. A descriptive research can be an extension or a pre-research stage to an exploratory form of research (Saunders & Lewis, 2012).

Explanatory research: It has the goal of analysing the causes and relationships. The explanatory research is about finding and explaining the relationships between the variables within the research area (Saunders & Lewis, 2012). The explanatory form of studies go beyond description and attempts to explain the reasons for the phenomenon which descriptive study only observes. Through explanatory research, the researchers uses theories or hypothesis to represent the forces that cause certain phenomenon.

This research work is exploratory and partial explanatory. It is because, the domain area of digital forensics is first explored by asking the research questions about the types of tools and methods available to gather digital evidences. The research is also characterized as partial explanatory because it attempts to explain as to how the digital
forensics tools/methods help in ensuring integrity, authenticity and accessibility of data for the digital collections hosted in Cloud platforms. The empirical investigations are conducted using selected digital forensics tools to highlight the application of digital forensics to digital asset management at the memory institutions or ALM.

3.2. Research Approach
The selection of research approach is another important aspect of conducting the research. The research approach focusses on the ways in which the research work has been approached (i.e. deductive vs. inductive, qualitative vs. quantitative etc.). The key characteristic of quantitative research is the selectivity and distance to the object of research whereas qualitative approach is subjective in nature and is characterized by the nearness to the object of research (Zikmund, 2000).

The qualitative approach focusses on the processes and meanings which cannot be measured in terms of quantity, amount, intensity and frequency. The qualitative approach provides a deeper understanding of the phenomenon within its context (Guba & Lincoln, 1994). The current research is about the applications of “digital forensics” to digitized collections and how it helps in ensuring integrity, authenticity and accessibility. Hence, the research context qualifies for qualitative approach with focus on relationships between digital forensics and how it applies to digitization on cloud.

The research design is another entity of research methodology. “A research design provides a framework for the collection and analysis of data. A choice of research design reflects decisions about the priority being given to a range of dimensions of the research process” (Bryman, 2008, p. 31). The five prominent research designs frequently being mentioned are (Bryman, 2008): experimental and related designs (such as the quasi-experiment); cross-sectional design, the most common form of which is survey research; longitudinal design and its various forms, such as the panel study and the cohort study; case-study design; and comparative design. The research study requires experiments to be conducted using digital forensics software to prove the concept of ensuring integrity, authenticity and accessibility of digital data.

3.3. Research Strategy
The research strategy chalks out the plan of action for conducting the research. It aims to find out the possible answers to the research questions. The research strategy defines the miscellaneous types of methods and tools that are related to achieve the
research goals. Generally there are two types of research strategies used for collecting data: quantitative and qualitative methods (Ghauri, Gronhaug, & Kristianslund, 1995).

The quantitative and qualitative methods can be taken to form two distinctive clusters of research strategy and both of these constitutes different approaches to social investigation and carry with them important epistemology and ontological considerations (Bryman, 2008).

The choice of research strategy depends on what the problem looks like, what questions the problem leads to, and what end-result is desirable (Merriam, 1994). There are five different types of research strategies: survey, archival analysis, case study, experiment and history (Yin, 1994). The choice of research strategy depends on several factors such as, the type of research questions, the need for control over behavioural events, and the degree of focus on contemporary vs historical events (Yin, 2003). The following table summarizes different research strategies (Yin, 2003).

<table>
<thead>
<tr>
<th>Research strategy</th>
<th>Form of research question</th>
<th>Requires control of behavioural events?</th>
<th>Focuses on contemporary events?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment</strong></td>
<td>how, why?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Survey</strong></td>
<td>who, what, where, how many, how much?</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Archival analysis</strong></td>
<td>who, what, where, how many, how much?</td>
<td>No</td>
<td>Yes/No</td>
</tr>
<tr>
<td><strong>History</strong></td>
<td>how, why?</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Case study</strong></td>
<td>how, why?</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The research for this thesis started by conducting secondary research and gathering some background information about digital forensics and cloud computing, subsequently the research questions were framed and the research strategy was devised. The research strategy chosen in this research work is mainly due to the contextual limitation of not finding the relevant case study organization and also the research topic not relevant for historic or archival analysis due to relatively new research area. However, the relevant digital forensic tools were experimented in the pilot setting to prove the applicability of digital forensics to digitization in cloud domain.
3.4. Data collection

The preparation of data collection can be complex and difficult (Yin R., 1994). If it is not handled properly, all the research outcome can be jeopardized. The data in the social sciences are gathered in either formal or informal settings and involve verbal or nonverbal responses (Nachmias & Nachmias, 1994). The four general forms of data collection may be elaborated as: observation methods, survey research, secondary data analysis and qualitative research. Every data collection method has some of the advantages and the limitations.

The qualitative researchers usually rely on four methods for collection information (Marshall & Rossman, 1999): (a) participation (b) observation (c) In-depth interviewing, and (d) analysing documents. There are basically six sources of evidence for collecting qualitative data: empirical documentation, archival records, interviews, direct observation, participant observation and physical artefacts (Yin, 2003).

The participant observation method is typically both an overall approach to inquiry and a data gathering. It is an essential element of all the qualitative studies and forces consideration of the role or stance of the researcher as a participant observer (Marshall & Rossman, 1999). The observation method of data collection are suitable for investigating the research area that researchers can observe directly. It covers systematic recording of events, behaviors, and artifacts in the social setting. It is a fundamental method of all the qualitative inquiry which is often used to discover complex interactions in the social settings (Marshall & Rossman, 1999). The analysis of documents often entails a specialized analytic approach called content analysis. It is an unobtrusive method which is often utilized as content analysis or archival analysis to gather records of a society, community, or organization to supplement other qualitative methods. The key strength of this method is that it is unobtrusive and nonreactive. It can be conductive without disturbing the social setting. A potential weakness is the span of inferential reasoning (Marshall & Rossman, 1999).

This data collection methods selected for this research study were mainly secondary data analysis (analysis of documents) and qualitative methods. The data was gathered from various sources such as EC Council (for digital forensics methods/tools), relevant bodies (law enforcement agencies), research bodies (NIST) and industry peers. The relevant information was also gathered on the state-of-the-art technologies related to the research work such as cloud computing models and frameworks.
3.5. Data analysis
The main aim of data analysis is to treat the evidence fairly, to produce compelling analytical conclusions, and to rule out alternative interpretations (Yin R., 1994). It involves examining, categorizing, tabulating, or otherwise recombining the collected data. Therefore, before even starting data analysis, a researcher can choose one of the two analytical strategies i.e. either relying on theoretical proposition or developing a case description. The qualitative analysis methodology can be used to gain an understanding of the underlying reasons and motivation to provide insights to the problem and generating ideas for further research.

There is no standardized approach for analysing qualitative data; rather the analysis most likely begins to an extent during data collection (Saunders & Lewis, 2012). The standardized data analysis process usually follows four step approach to process, to inspect, to transform and to model data. The four step data analysis process includes: 1) data preparation, 2) data exploration including coding and analysis, 3) data reducing or filtering which can be done with data exploration, 4) interpretation including analysis of findings and thematic analysis (Hesse-Biber, Sharlene, & Leavy, 2008).

This research study is going to apply qualitative analysis methods which employs analysis of data or content analysis (Kakooza, 1996). Firstly, the emerging issues or themes will be identified related to digital forensics, cloud computing and digitization. The exploration involves systematic identification of characteristics of the findings and the related secondary data to support positions and provide necessary authenticity to the research findings for drawing conclusions. The analysis of experimentation results for digital forensics software to prove the integrity, authenticity and accessibility parameters will provide further justification to the research findings / conclusions.

3.6. Research quality criteria
The research work requires two important underlying concepts to be considered, validity and reliability. These two criteria are extremely important to verify the adequate applicability of the selected research methods to the research work. The criteria of validity is about the measurement and the ability to measure what the research is supposed to measure. The criteria of reliability is a measure of how stable and consistent the measurements are, if the research would find the same results when repeating the research. This research work have more than one sources of evidence when collecting data, documentation and industry discussions with peers.
The documentation included various policy documents, best practice, laws, technical documents, academic journals, books etc. The analysis of documentation have built the theoretical foundation and provided the necessary background for conducting this research work. The various forensics processes, models, architecture, and frameworks were explored to prove the point of digital forensics applications to digitization on cloud. The discussions with peers have provided an opportunity to validate findings. The experimentation using digital forensics software have provided another possibility of validating the findings and prove the applicability of digital forensics to cloud.

3.7. Summary
The research methodology presented in this chapter has highlighted the research approaches, research strategy and the research design to conduct this research work. It was mentioned that the research is mainly exploratory with partially explanatory. The research strategy and research design has been elaborated to define data collection methods. The chapter ended with the selection of data analysis methods, and the quality criteria of validity and reliability to evaluate the methodology of research.
Chapter 4
Analysis and Findings

This chapter covers the analysis of the research problems based on the defined research strategy and the chosen methodology for conducting research. The chapter presents the analytical findings for the research questions through pragmatic exploration on the topic of digital forensics. Wherever required appropriate digital forensics tools were utilized for the empirical investigations. Finally, the findings of analysis are presented to highlight the outcome of the research work.

4.1. Introduction
The growing cyber threats and the increasing value of digital data are the major reasons for the growth of digital forensics. It has emerged as an essential source of growth in tools and methods for facilitating management of digital assets, specifically for protecting and investigating digital evidence from the past. The institutional repositories and the custodians of memory institutions can benefit from forensics in addressing digital authenticity, accountability, integrity and accessibility of digital data. The digital data for personal information must also be handled securely while demonstrably protecting its evidential value. The tools and methods in digital forensics makes it possible to establish chain of custody for provenance, employ write protection for data capture and transfer, identify privacy issues and detect digital forgery and data manipulation. Digital forensics can help extract and mine relevant metadata, enable efficient indexing and searching, and facilitate audit control and access privileges. The increasing sophistication in digital forensics are offering new tools and methods for effective automation in the handling of higher volume of personal digital information. By applying right framework and policies, the digital forensics could offer endless possibilities of both protecting the digital data for authenticity, integrity and accessibility as well as to help capture digital evidences for presentation in the court of law. This research work provides a broad overview of digital forensics applications to the digital collections hosted in the cloud environment, with some tools, resources and framework that may benefit memory institutions in the curation of ALM digital resources.

4.2. Long-term retention of digital data
The field of Digital preservation concerns with the management of digital materials of time to ensure sustainable access. It may be defined as: “the series of managed activities necessary to ensure continued access to digital materials for as long as
necessary, beyond the limits of media failure or technological and organisational change” (Beagrie N., 2001). The emphasis of the definition is on ensuring resilience and perceptibility of digital assets for long-term accessibility. This definition highlights both technical and organizational challenges in maintaining digital materials over time. The major risks to sustainability lies in the growth of technology, erosion of evidential quality, uncertainty for authenticity and completeness of content.

Tab. 4 Industry requirements for long-term retention of data

<table>
<thead>
<tr>
<th>Industry</th>
<th>Typical data retention requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>Patient lifetime</td>
</tr>
<tr>
<td>Pharma</td>
<td>75-100 years for R&amp;D records; applications</td>
</tr>
<tr>
<td>Energy</td>
<td>Oil field: lifetime</td>
</tr>
<tr>
<td></td>
<td>&gt;50 years of sub-surface exploration data</td>
</tr>
<tr>
<td>Government</td>
<td>&gt;100 years</td>
</tr>
<tr>
<td>Finance</td>
<td>Lifetime of the life insurance policy, plus 6-10 years</td>
</tr>
<tr>
<td>Aerospace</td>
<td>&gt;30 years for design records (lifetime of aircraft)</td>
</tr>
</tbody>
</table>


The industry requirements for long-term retention of data puts pressure on IT to maintain digital data for long-term while checking for technology obsolescence and data destruction. The digital repositories for maintaining digital objects are very diverse in terms of their organization and content. Critically, the historic and scholarly value of almost any kind of digital objects depends on their evidential properties – reflected in, for example, the nature of the embedded metadata.

To support the long-retention of digital data, some of the repositories have turned to digital forensics, to offer solutions for the effective curation of digital repositories, and for the automated management and analysis of collections (Kirschenbaum, 2008; John, 2008). The forensics tools are useful across range of preservation functions for organizing digital collections, mainly with complex digital objects such as social media archives, electronic records/documents, applications, web archives etc. The forensics
tools are very useful in context of digital preservation to understand the consequences of file format migration, reliability of emulation software, data recovery from crashed devices and evaluating the damage from cyber-attacks or espionage. Conversely, the preservation methodologies also contribute significantly to digital forensics practice.

4.4. Implementation of Cloud-based digital archiving

The digital preservation is an important function for any industry requiring a long-term data retention plan. There is an increasing demand for data storage solutions for both born-digital archives and digitized materials at the organizations. The growing volumes of digital materials requiring preservation come from many different sources including those created or acquired in digital form or via digitizing of existing physical collections. The varied types of digital collections may have different characteristics and preservation needs. The digital archives are becoming more widespread in the organizations, and the OAIS reference model (ISO 14721) provide a set of concepts and definitions assisting implementation across the industry and facilitate the specifications of archives and digital preservation systems (Beagrie 2014).

With the growth in technology and the growing migration of organizations from physical infrastructure to virtualization, similar trend is also being witnessed in the digital preservation, where cloud may be a component of the required solutions enabling wider participation and collaboration. Nearly all the providers of cloud infrastructure services are offering cloud-based storage solutions. The archives or digital preservation systems, typically have additional requirements beyond accessibility of storage space to host files. The specific requirements may include concerns for data protection, processing of PII (personal identifiable information), risk attitude, value of data loss, and long-term accessibility of data (20-50 years) etc.

The generic CSPs, such as Google, Amazon, Microsoft and others, do not typically address specific requirements for archiving / preservation within their basic offerings. However, there are number of other specialized cloud providers offering value added services with an additional layer of processes and procedures on top of generic services to build customized digital archiving solutions on Cloud environment. It is often assumed by the customers that the cloud solutions only run on large data centers accessed online over the internet. It is one of the model (public cloud), but there exists other models which allow installation locally and access over private networks (private cloud), or even flexibility to combine two models with local storage (hybrid cloud). The
Another open source solution called Archivematica offers cloud based hosting solution for digital archiving. The cloud based archiving environment offers various benefits to the organizations in terms of improved enterprise architecture, achieving compliance, mitigating risks and increased IT efficiency and costs. But, the potential benefits from solutions also bring significant issues in terms of information security and legal requirements. One of the major vulnerabilities of Cloud solutions is Security, particularly for the sensitive, commercial and/or PII (personally identifiable information). The large CSPs put huge efforts in ensuring security, with physical security of the data centers and regular audits (internal/external) to ensure compliance to ISO 27001 Information Security and ISO 22301 Business Continuity requirements.
Some of the cloud initiatives also implement specific standards such as ISO 27799 to provide additional specificity in the domain such as storing patient data.

The term ‘cloud’ encompasses wide range of implementation models for archival storage. There are possibilities to gain deeper understanding about the implementation from the archives who have already piloted or moved to use cloud storage. Several archives have been able to address the biggest concerns over cloud services and successfully integrate cloud storage in their preservation activities. Few of the pros and cons of implementing cloud for digital archiving (Beagrie & Charlesworth, 2014):

### Tab. 6 Pros and Cons of Cloud for Digital Archiving

<table>
<thead>
<tr>
<th><strong>Pros</strong></th>
<th><strong>Cons</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cloud services can provide easy, automated replication to multiple locations and access to managed digital storage and integrity checking.</strong></td>
<td><strong>Cloud is designed for flexibility and rapid change.</strong> Service contracts for cloud storage need careful management. Data held in cloud archives is expected to be preserved and accessible far beyond the commercial lifespan of any technology or CSP.</td>
</tr>
<tr>
<td><strong>Archives can add access to dedicated tools, procedures, workflow and service agreements, tailored for digital preservation requirements via specialist vendors;</strong></td>
<td><strong>Cloud can be cheaper, but the organizations require different way to manage their budgets especially in terms of vendor management and training/re-training costs.</strong></td>
</tr>
<tr>
<td><strong>Potential cost savings from easier procurement and economies of scale;</strong></td>
<td><strong>Public cloud services often bill monthly for the capacity that has actually been consumed. It can be difficult to predict the usage / amount of data uploaded, downloaded or deleted (Vendors often provide annual subscription based on volume).</strong></td>
</tr>
<tr>
<td><strong>Flexibility of cloud allows relatively rapid and low-cost testing and piloting of providers;</strong></td>
<td><strong>Cloud storage is a form of outsourcing, thus it is important that archives exercise diligence in controlling risks. It is recommended to ensure any legal requirements (contracts, agreements) for the management, preservation and access by third party.</strong></td>
</tr>
<tr>
<td><strong>Greater flexibility and more options in deployment of cloud services and therefore greater relevance to archives compared to earlier years.</strong></td>
<td><strong>Cloud services will require archives to consider copyright related matters; who owns it; if license is required; what permissions the CSP will need; Whether CSP is able to use data; which party will own rights in any data or works created from the original data;</strong></td>
</tr>
<tr>
<td><strong>Exit strategies can be put in place to address archival concerns over provider stability and longevity or other risks e.g. synchronizing content across two CSPs with local internal storage; or agreeing an escrow copy held independently by trusted third-party;</strong></td>
<td><strong>Cloud services may raise data security issues, where the relevant data is ‘personal data’, include determining responsibility for securing data and audit of CSPs;</strong></td>
</tr>
<tr>
<td><strong>Explicit provision for pre-defined exit strategies, testing, monitoring &amp; audit procedures.</strong></td>
<td><strong>Require well-defined legal relationship between an archive and CSP (e.g. SLA, terms of service contracts). Many CSP have standard SLA / contracts for commodity pricing;</strong></td>
</tr>
</tbody>
</table>
The implementation of cloud-based digitization or archiving solution is not so difficult if the organization has a well-defined implementation plan. A step by step approach to implement cloud may be summarized in the following steps:

1. **Step 1: Establish the needs** – Identify essential ‘must have’ and ‘could have’ requirements. The process may be iterative particularly if the archive is new. A careful planning of pre-implementation phase is required, with the involvement of staff and departments to validate requirements, capabilities and budget.

2. **Step 2: Service Options** – Cloud computing offers range of different scenarios, from IaaS to SaaS and HaaS, the public cloud is the most widely used for commercial services hosted in large data centers around the world whereas the private cloud is often used by the organizations having significant investment in data centers, equipment and manpower. There is a growing trend of offering cloud-hosted or locally-hosted solutions from CSPs for archival storage and other preservation functions. Amazon, for example offers a complete set of cloud storage services for archiving. The organizations can choose Amazon Glacier for affordable, non-time sensitive cloud storage, or Amazon S3 for faster storage, depending on organizational needs and budget (Amazon Inc., 2017).

3. **Step 3: Service providers** – There are many different ways in which archives may choose to deploy cloud storage, from simple replacement of local physical disk with cloud-based storage through to more complex arrangements with fault-tolerant workflows and policies. There exists two categories of cloud storage service providers: Generalist (Google, Amazon etc.) or Specialist (Arkivum, Archivmatica, Preservica etc.). The generalist are cheaper and easily accessible globally but they may not always meet specific archival requirements such as ingest, processing, long-term storage or maintenance of data. The specialist are little bit more expensive and often operate with longer subscription period. The specialist providers are often intermediaries who take generalist infrastructure (such as Amazon, Microsoft or Google) and layer the archival workflows and processes on top to create solutions for archival sector (e.g. a US-based Internet Archive, offers a well-known service called Archive-It). It is explicitly designed to support the archiving of web sites (Archive-IT, 2017).

4. **Step 4: Preparation of business case** – It is an important step before implementation to create a business case and the justification in terms of costs
and sustainability of data. It is always recommended to follow the template and the process being followed in the specific organization.

![Step-by-Step Cloud-based Archiving Implementation Lifecycle](image)

**Fig. 11 Step-by-Step Cloud-based Archiving Implementation**

5. **Step 5: Implementation** - It is recommended to follow step by step approach for the implementation starting with pilots and proof of concepts. Start out small and progressively scale up to more critical areas of content. A digital archive may be diverse with different requirements e.g. data encryption while uploading, secure management of security keys, encryption of data while on cloud, and risk management process to check mis-management. It is important to check that the provider offers regularly audited information security with compliant procedures as those documented in the ISO 27001:2013 ISMS requirements.

6. **Step 6: Post-implementation**: It is important to capture and share lessons learnt from the cloud-based digital archiving implementation. It is recommended to use OAIS model (ISO 14721) to facilitate discussion as well as new improvements in the specifications of archives and digital preservation systems. Furthermore, it is encouraged to get certification, such as ISO 16363 for audit and certification of trustworthy digital repositories and Data Seal of Approval to provide formal standards of accreditation of digital archives (Beagrie 2014).

The abovementioned step-by-step implementation lifecycle explains the typical process of implementation of cloud-based archiving solution. Since, the security concerns (such as privacy, confidentiality and integrity) are highlighted as the most critical for cloud based implementation, hence it is extremely important to consider
application of digital forensics to the cloud-based implementation. This thesis work explores the specific tools and methods of digital forensics to the collections hosted in cloud environment. The next section explores digital forensics in context of archiving.

4.3. Digital forensics in Archiving
The field of digital forensics has applications in both crime investigations and towards facilitating digital preservation and curation, specifically for protecting and investigating the historic evidences. The digital repositories are very complex with sheer diversity of objects (such as file types, formats and applications) and intricate structural relationships posing significant challenges to preservation. Mostly, the digital archives may have almost anything hosted such as drawings, 3D designs, emails, draft reports or datasets. The forensic procedures may be applied to such collections to ensure the sustained preservation of the digital objects for long-term retention.

There are 3 basic principles in digital forensics: 1) the evidence is acquired without altering it; 2) it is demonstrable; 3) the analysis is conducted in an accountable and repeatable way. The digital forensics processes and tools are designed to ensure compliance to these requirements. The application of digital forensics is applicable holistically to the curatorial and preservation lifecycle. The representative steps of a forensics lifecycle for a hard drive may be summarized as follows (Beagrie, 2012):

1. Remove the disk from the originator’s computer;
2. Attach the collected disk to the forensics computer via a write-blocker to prevent the disk from being altered;
3. Capture a mirror image of the disk representing entire content of hard disk;
4. Create a cryptographic hash for all the digital objects and the disk itself;
5. Test for malware and viruses;
6. View files using forensics hardware / software;
7. View the metadata such as date and time of file creation;
8. Extract the metadata and record in metadata in cataloguing system;
9. Identify and bookmark privacy concerns e.g. digital personal information such as credit card information, social security number etc.
10. Export replication of original objects and of disk image and examine in an emulator;
11. Convert digital replication to interoperable files;
12. Comply with digital preservation guidelines;
13. Analyse metadata and content of objects to create exploratory visualization
14. Save log files of the forensics examination process;
15. Create a forensic report documentation for capture and analysis by curator

The aforementioned steps highlights the entire lifecycle of the steps taken to conduct the forensics examination of a hard drive. There are varied software and hardware tools available to support the specific steps in the forensics lifecycle. The technologies of forensics vary greatly in terms of their capability, costs and complexity. Some of the equipment are expensive whereas others are available for free. The techniques such as anonymization, encryption, selective redaction, malware sandbox containment and other mechanisms may be required to assure that privacy is fully protected and inadvertent leakage of information may be prevented. The use of computers in open WIFI network, portable devices and shareable cloud services all harbour significant PII (personal identifiable information) and consequently raise issues of privacy.

![Forensic tools in the archive](image)

**Fig. 12 Digital forensics in context of OAIS model**

The digital forensics experts and the custodians of digital information share a common need to monitor and understand as to how the technology is used to create, store, and manage the digital information. Additionally, they also need to manage the information conformance to the relevant standards and best practices applicable for the specific industry. The application of forensic techniques could help handling of digital
information from remote computers, to networks, mobile devices, flash media, virtual machines, cloud services and encrypted devices.

The digital forensics and preservation methodologies retrospectively interpret events represented on the digital devices, and to react rapidly to the transforming landscape by the rapid institutionalization of policies, procedures and solutions (Beagrie, 2012). The principles in digital forensics have important practical implications in curatorial universe. The functionalities such as file viewing, analysis of file signature, identification of known files (by means of hashing), data extraction, file export, searching, indexing, bookmarking, timeline visualization, logging, and reporting, all of which are to be undertaken according to forensically sound principles (Beagrie, 2014).

This research work examines the applications of methods/tools of digital forensics to cloud based archiving. The specific applications for ensuring authenticity, integrity and accessibility would be tested by applying a process-approach. The analysis of findings would present available methods/tools, process, and test results in the next section.

4.5. Analysis of research questions
The findings of the analysis of research questions is presented in this section. The data collected in the study is thoroughly analysed to elaborate the research findings. The first research question elaborates the application of digital forensics tools and methods with explanations of the specific software being used in forensics applications. The second research question focusses on the process dimension and elaborates the ‘how to’ of digital forensics. Finally, the third research question examines the process-approach and how it may be applied to the digital collections.

4.5.1. What are the different methods, tools and techniques in digital forensics?
The application of forensics technology makes it possible to: establish chain of custody for provenance, employ write protection for capture/transfer, identify privacy issues and detect manipulation and forgery. Digital forensics can extract and mine metadata, enable indexing and searching, facilitate audit control and access privileges. It has emerged as a key source of tools and methods for facilitating digital preservation specifically for protecting and investigating evidences from the past. The custodians of memory institutions can benefit from digital forensics in addressing authenticity, accountability and accessibility. The digital materials and PII (personal identifiable
information) must be handled securely respecting available rules (Tab. 7) and standards while demonstrably protecting the evidential value.

**Tab. 7 Rules for digital forensics investigations**

| 1. Limited access and examination of the digital evidence | 6. Evidence should be strictly related to the incident |
| 2. Record changes made to the evidence files | 7. The evidence should comply with the jurisdiction standards |
| 3. Create chain of custody document | 8. Document the procedures applied to the evidence |
| 4. Set standards for investigating the evidence and comply it | 9. Securely store the evidence |
| 5. Hire experts for analysis of evidence | 10. Use recognized tools for analysis |

The forensics investigation process is a collection of many processes starting from incident response to analysis of the crime scene, gathering evidence to its analysis, and from documenting to reporting. Prior to forensics investigation start, it is important to understand the scope of investigation, what evidence is needed to be acquired and for what there is an explicit authority to be acquired. After understanding the scope, the next task is to understand the type of investigation to be conducted.

**Live acquisition:** Open devices – What is in memory OR if the network state needed?

**Static acquisition:** Open devices – What is the status of open files and/or programs.

![Fig. 13 Forensics investigation process](image)

The first responder has to identify and select the type of investigation and based on that it has to be determined if it is needed to keep the system up and running to do live acquisition or only hard drive and other storage device are needed for static acquisition. The step of search and seizure is followed by storage of disk drive, USB disk or any
type of memory card or PC. There may also be a need for secure storage of other
digital artefacts to store image files of evidence that cannot be tampered with. There
must be a chain of evidence and evidence verification (hash values) be maintained to
secure the evidence before it is presented in court as a testimony. There exists several
investigation tools which are used for data acquisition and data analysis such as:

**Tab. 8 Methods and tools in Digital forensics**

<table>
<thead>
<tr>
<th>Forensics method</th>
<th>Techniques of investigation</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>File system analysis</td>
<td>File system analysis</td>
<td>X-Ways, Hex Editor</td>
</tr>
<tr>
<td></td>
<td>File carving</td>
<td>Sleuth Kit (+Autopsy)</td>
</tr>
<tr>
<td>Data acquisition and duplication</td>
<td>Bit stream image and backup copy</td>
<td>ProDiscover forensics</td>
</tr>
<tr>
<td></td>
<td>Write protection and media sanitization</td>
<td>FTK Imager</td>
</tr>
<tr>
<td></td>
<td>Data acquisition formats (raw, proprietary, AFF)</td>
<td></td>
</tr>
<tr>
<td>Anti-forensics techniques / tools</td>
<td>File recovery and partition recovery</td>
<td>SANS Investigative forensics Toolkit</td>
</tr>
<tr>
<td></td>
<td>Password protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Steganography and Steganalysis</td>
<td></td>
</tr>
<tr>
<td>OS forensics</td>
<td>Memory and registry forensics</td>
<td>Process Explorer</td>
</tr>
<tr>
<td></td>
<td>Windows file analysis</td>
<td>ProDiscover forensics</td>
</tr>
<tr>
<td></td>
<td>Event log analysis</td>
<td></td>
</tr>
<tr>
<td>Metadata analysis</td>
<td>Metadata investigation for files / OS</td>
<td>Metashield analyser</td>
</tr>
<tr>
<td>Network forensics</td>
<td>Log analysis</td>
<td>GFI Event Manager</td>
</tr>
<tr>
<td>Web attacks</td>
<td>Server logs</td>
<td>Nagios Log Server</td>
</tr>
<tr>
<td>Cloud forensics</td>
<td>Log monitoring / Due diligence</td>
<td>UFED Cloud Analyzer</td>
</tr>
<tr>
<td></td>
<td>Data and system recovery</td>
<td></td>
</tr>
<tr>
<td>Malware forensics</td>
<td>Static and dynamic malware analysis</td>
<td>IDA Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Directory Monitor</td>
</tr>
<tr>
<td>Mail forensics</td>
<td>Email crimes and violations</td>
<td>MailXaminer</td>
</tr>
<tr>
<td>Mobile forensics</td>
<td>Mobile OS and forensics</td>
<td>Oxygen forensics</td>
</tr>
</tbody>
</table>
Forensic investigators must do everything possible to preserve the integrity of the digital evidence. Any mistakes in the process call the evidence into question and rendering it worthless. In case of accidental or intentional hiding, deleting or loss of data, the data recovery methods / tools are used to recover digital evidences. There are handful of tools available to facilitate recovery, few of them are listed below:

- **VirtualLab Data Recovery:** It supports NTFS, FAT32, Mac HFS, HFS+ and NFS file systems. It supports data recovery on devices such as memory card or USB drives. It has an ability to make sector-by-sector copy of a damaged drive.

- **EaseUS Data Recovery:** It is a complete recovery solution having 3 modules:
  - Complete Recovery: It is used to recover data from formatted / corrupted drive as a RAW.
  - Deleted File Recovery: It is used to recover data deleted from recycle bin.
  - Partition Recovery: It is used to recover data from deleted / hidden / lost hard drive partition.

- **Stellar Phoenix:** It is used to recover data from Windows PC hard drive, memory card and USB sticks. It can restore data from archives, databases, documents and different type of multimedia files.

The integrity of data is handled in every step of forensics investigation process such as the way in which data is seized, labelled, transported, copied, analysed and finally presented in the court. The significant efforts are made to guide the first responders as to how to collect evidence at the scene of investigation. But, the evidence handling continues throughout the lifecycle of the digital forensics, many of the procedures and stipulations strike a chord with archivists, curators and digital preservation experts.

The first research question has investigated the different methods, tools and techniques in digital forensics including the description of process (if any) to aid the forensics investigation. The findings have highlighted the availability of several proprietary and open source tools to support methodical investigations. The next section would present the ‘how to’ or ‘process-approach’ for implementing it.

**4.5.2. How can digital forensics be applied to capture digital evidence for ensuring authenticity, integrity and accessibility??**

The intersection between digital forensics and archives can be characterized as an intermediary zone between both fields of activity which symbolizes their mutual
relationships in practice. The three fundamental concepts of archiving can be advanced through the adoption of digital forensics methods and tools:

1. Provenance,
2. Original order and
3. Custody

The ‘provenance’ information of a document is actually a life history of it. It is characterized as one of the most important information for describing identification or origin or sources of the document in the digitized collection. According to the principle of provenance, a document or a records from a common origin or source should be managed together as an aggregate unit and should not be arbitrarily intermingled with records from other origins or sources (Lee C., 2012). Another closely related aspect to provenance is the principle of original order which reflects in organizing and managing of records in their order of arrangement.

The order of arrangement in the digital environment conveys information about the record-keeping, and additional layers of description can be elaborated to facilitation navigation and access. For the purpose of compliance, authenticity, integrity and admissibility, the record-keeping system must control, documentation and manage account of record for point of creation, use, re-use and destruction. The ‘original order’ may be understood through chain of custody which is a record held for all the digital materials in the collection to detail out the lines of responsibility, changes of states (file formats and version control) or location (filing system or storage) of the digital asset.

The digital assets stored in the digital preservation systems follows the archival view involving principles, policies, and preservation strategies designed to ensure the stabilization of records to extend its lifetime, protect accuracy and authenticity, and maintaining accessibility to the digital content (Duranti & Jansen, 2011). The preservation includes description as a key element which provides context, attributes, and relationships of the record aiding to develop and maintain the retrieval system.

The identity of a digital record in the collection comprises of distinct attributes that characterizes and differentiates it from other records. To prove the authenticity of a digital record, it is essential to establish the continuing identity (chain of custody) and prove integrity (Duranti & Jansen, 2011). The identity can allow forensics investigator to differentiate one record from another and to determine who, what, when, why as well as to establish the degree of granularity / versioning (i.e. draft, original or a copy).
The custodians of the digital materials can only preserve the records as trustworthy (i.e. authentic, accurate and reliable) at the time of collection and storage. It is therefore their responsibility to establish the identity of the record prior to acquisition and then to maintain the identity together with their integrity (MacNeil, 2004).

The digital forensics methods and tools are commonly being applied in digital archiving by treating disk images rather than individual files or directories as the basic unit of data acquisition. Furthermore, the use of write-blockers, full disk image creation, and data extraction associated with files are essential to ensure provenance, original order and chain of custody. The application of digital forensics methods to archiving are essential to ensure the sustainability of archives as the custodians of PII (personal identifiable information); the forensics tools that are commonly used to expose the sensitive information can be used to identify, flag, and restrict the access to it.

Evidence identification

Evidence collection

Evidence integration

Evidence preservation

Evidence analysis

Evidence presentation

Fig. 14 Process-approach to digital evidence handling

The interaction explained between digital archiving and digital forensics has highlighted authenticity, integrity, identity, and accessibility as the parameters to ensure the preservation of digital record. The specific forensics tools such as HashCalc, HEX editor and X-Ways forensics offers the functionalities to ensure integrity, accessibility and authenticity of digital record. The digital evidence handling in the forensics investigation process (Fig. 14) must apply the specified parameters to ensure preservation of digital evidence prior to presentation in the court.
4.5.3. Which digital forensics methods and tools be used to capture digital evidence for digitized collections in Cloud environment?

The Cloud paradigm is offering tremendous benefits by providing highly scalable infrastructure, pay-per-use model, and low cost of on-demand computing. Cloud is fast growing in the organizations. But, the security and trustworthiness is still number one concern for the cloud infrastructure. The malicious individuals can simply manipulate attacks from nodes / hosts inside the cloud. A compilation of Top 10 threats and attacks in cloud computing environment could be summarized as follows (EC Council, 2017):

<table>
<thead>
<tr>
<th>Threats Top 10</th>
<th>Attacks Top 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data breach / loss</td>
<td>Service hijacking using social engineering</td>
</tr>
<tr>
<td>Abuse of cloud services</td>
<td>Session hijacking using XSS attacks</td>
</tr>
<tr>
<td>Shared technology issues</td>
<td>DNS attacks</td>
</tr>
<tr>
<td>Insecure interfaces and API</td>
<td>SQL injection attacks</td>
</tr>
<tr>
<td>Hardware failure</td>
<td>Wrapping attacks</td>
</tr>
<tr>
<td>Conflict between client hardening procedures &amp; cloud environment</td>
<td>Service hijacking using network sniffing</td>
</tr>
<tr>
<td>Privilege escalation</td>
<td>Session hijacking using session riding</td>
</tr>
<tr>
<td>Authentication attacks</td>
<td>Cross guest VM breaches</td>
</tr>
<tr>
<td>VM level attacks</td>
<td>Cryptanalysis attacks</td>
</tr>
<tr>
<td>Improper data handling &amp; disposal</td>
<td>DoS and DDoS attacks</td>
</tr>
</tbody>
</table>

The complexity of cloud attacks and the threat scenario makes the cloud forensics process complex. The storage are mostly not local in cloud (except for private cloud). Each server or host may have files for many users. Therefore, it is not easy to seize the server or host from a data center without violating the privacy of other users. Moreover, it is also difficult for the law enforcement agencies to seize suspect’s digital devices in order to get access to the digital evidence.

It was far easy for the forensic investigators in the traditional digital forensics to get physical access and full control of evidences (e.g. hard disk, logs etc.). Unfortunately, in cloud computing, there exists a shared responsibility model which works according to the hosting model selected by the customer. There are different levels of control and responsibilities between the CSP and the customer for the three different service models (i.e. IaaS, PaaS, and SaaS). The Cloud users have highest control in IaaS and least control in SaaS. The lack of physical access and the absence of control over the system makes it difficult to perform evidence acquisition in the cloud environment. The source of evidence in cloud environment is ubiquitous and the linking to the source is
complicated. Furthermore, the investigators must have to determine the computational structure, attributes and integrity of data to perform the investigation which is not easy. Also, the investigators have to keep the stability of evidence and visualize it.

There are many ways to conduct forensics investigation in cloud environment. The investigations may be similar to the traditional forensics investigation. In the first way, the cloud may be considered as a tool of the crime, whereas in the second way, the cloud hosts a service as a target of the crime. There are three major aspects to be considered in the forensics investigation for cloud environment:

1. Nature of the crime determines the type of the system (alive or dead)
2. Determination of what actually took place in the cloud environment
3. Availability of secure channel to collect evidences over the cloud

The digital forensics has two scenarios for digital evidence acquisition (i.e. live system / dead system). In the dead system, the investigators only analyse hard disk images whereas more evidences may be captured from running processes in the live system. Each node (i.e. server, router, and switch) in the cloud environment can be a source of forensic data. The CSPs provides only limited access of evidence set for forensics analysis to the customer. The trickiest thing in the cloud implementation is to localize and identify the instances of data (e.g. VMs or applications used by users may migrate number of times between various physical servers with or without any record-keeping). Also, the access to forensics data and digital evidence may be limited for the customer by degree of implemented technology (e.g. storage space), price, virtualization multi-tenancy, privacy and other conditions connected with the CSP infrastructure. One of the major advantage of implementing digital forensics in cloud environment over the traditional environment is that the digital forensics in cloud is considered a live system which gives the investigator more files, connection and services to be acquired and investigated. The CSPs are obliged to log all the activities and have forensic support for all the services offered and used by the customer. The different service models offer basic terms of cloud forensics for users (Dobrosavljević & Veinović, 2015):

1. **SaaS:** The customization options and preferences of the customer are limited in the SaaS model. The users do not have control over physical infrastructure such as network, servers and the source code of applications. The restricted access limit customers to conduct analysis of log files and perform forensics. SaaS model requires detailed implementation of application logs for each
application hosted in cloud. It is recommended that the CSPs and the customers must always agree on the details about the forensics in the SLAs.

2. **PaaS:** The entire development environment and all the source codes never leaves the development platform in PaaS model. The customer has an option to install any forensics tool within their own application. The remote logging server can be installed and automatic logging options can be implemented to create a single repository of all logs and events. In PaaS model, the user is responsible for the functionality of the application, whereas the CSP guarantees that the application is available and operational. The customer needs to create the responsibility boundaries between users and the CSP wherever there is a need for forensic data. These responsibilities must be documented in the SLA. The SLAs must detail procedures when accessing notification logs, identification logs, preservation logs and all the potential evidences.

3. **IaaS:** The most number of options for configuration along with logging features and high level of control are offered in IaaS model. The user controls most of the components of the system including log sources. The users and CSPs must document SLA for forensics data collection and logging.

![Fig. 15 High level Cloud forensics process](https://example.com/fig15)

*Source: Cloud Security Alliance (2015)*

The responsibilities between the customer and CSPs including the responsibilities for forensics can be understood from the cloud forensics process (in Fig. 15). The detailed process steps, challenges and solutions by service models are presented in Tab. 10

**Tab. 10 Cloud forensics process phases and solutions**
<table>
<thead>
<tr>
<th>Phase</th>
<th>Challenges</th>
<th>Solution</th>
<th>IaaS</th>
<th>SaaS</th>
<th>PaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence identification</td>
<td>To find out relationship among many evidence files</td>
<td>Metadata analysis to identify file association / relationships</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence collection</td>
<td>Identify and rank remote malicious hosts authentically</td>
<td>Select hosts based on probability to minimize costs</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evidence collection</td>
<td>Identifying cloud attacks and applying suitable forensics</td>
<td>Identify syslog relevant to attack and apply forensics accordingly</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence collection</td>
<td>Robust SLAs</td>
<td>Robust SLA to ensure IS</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence collection</td>
<td>Collecting consistent evidences from VMs</td>
<td>Efficient log monitoring for VMs</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evidence collection</td>
<td>Identifying as well as localizing SSL/TSL attacks</td>
<td>Implement a tool to identify MitM attack on SSL/TLS</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence integration</td>
<td>Suitable monitoring system based on virtualization</td>
<td>Implement tool to monitor live system</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evidence integration</td>
<td>Easy access to outsourced data collected for forensic</td>
<td>Leaked Access Credential Tracing for evidence collection</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence integration</td>
<td>Designing suitable model for verifying data integrity</td>
<td>Implement fault-tolerant and &amp; integrity verification scheme</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evidence integration</td>
<td>Verifying integrity for secure forensic analysis</td>
<td>Implement fixed rate cloud model for integrity verification</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Evidence integration</td>
<td>Data integration with the help of auditability</td>
<td>Implement evidence integrity model over tree structure</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Evidence preservation</td>
<td>Ensuring protection of evidences from tampering</td>
<td>Division of cloud architecture &amp; movement of cloud instances</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence preservation</td>
<td>Ensuring secure logging and management in cloud</td>
<td>Ensure privacy &amp; confidentiality of log data through encryption</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Evidence analysis</td>
<td>Suitable &amp; sound forensics tool for Open Stack</td>
<td>Incorporate FROST instead of guest VM for cloud management</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence analysis</td>
<td>Implementation of anti-phishing techniques</td>
<td>Introduce offline phishing identification system</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence analysis</td>
<td>Memory forensics in the cloud environment</td>
<td>Implement memory forensics techniques for analysing VM</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evidence presentation</td>
<td>Increasing trust amongst cloud data centers</td>
<td>Implement solutions to manage trust across cloud data centers</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Evidence presentation</td>
<td>Chain of custody for the presentation in court</td>
<td>Ensure chain of custody included, (evidence collector, collection procedure, storage, accessibility)</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
As the danger of data leaks / data loss becomes more critical for business continuity, it is becoming important for the customers to define SLAs with CSPs to include the option for forensic investigations. The CSPs must define and/or agree on the procedures that CSPs needs to follow in case of forensic data acquisition, including methodology for evidence identification and preservation against tampering / damage.

![Figure 16: Implementation of ISO standards in Cloud forensics](source: ISO (2015))

The most critical aspect of evidence handling is the preservation of evidences, because they are notoriously fragile, and must be handled with utmost attention. The storage must have access control policies to prevent or protect the potential evidence from tampering, as well as kept with the appropriate storage conditions. The various ISO standards such as ISO 27041 / 42 / 43 have the standard process, procedures and approach defined to implement digital forensics in Cloud environment (Fig. 16).

**Digital forensics applications in the Cloud environment of Google Drive:**

The Google Drive client is first installed with configuration stored locally and synced to the cloud storage online.

1. Google Drive client is installed:
   C:\Program Files\Google\Drive
2. Client configuration is stored in the user profile:
   C:\Users\<username>\AppData\Local\Google\Drive
3. Default folder used for syncing is:
   C:\Users\<username>\Google Drive
4. By default the user folder is empty
5. During installation different keys and values are created inside the registry. Investigators can gather information from registry on:
   - Installed version
   - User folder

   SOFTWARE\Microsoft\Windows\CurrentVersion\Installer\Folders\  
   SOFTWARE\Google\Drive  
   NTUSER\Software\Microsoft\Windows\CurrentVersion\Run\GoogleDriveSync  
   NTUSER\Software\Classes

6. Configuration files are saved inside the installation folder in the user profile. Executable and libraries are stored in the bin sub-folder.

   C:\Users\<username>\Desktop\Google Drive.lnk  
   C:\ProgramData\Microsoft\Windows\Start Menu\Programs\ Google Drive\Google Drive.lnk

7. The Snapshot.db is a SQLITE3 DB containing information about local and cloud entry.
   - Cloud_entry table
   - File name
   - Created (UNIX Timestamp)
   - Modified (UNIX Timestamp)
   - URL
   - Checksum (MD5 hash)
   - Size
   - Shared

8. Sync_log.log is a log file containing information about the client sync session. Investigators can gather information about:
   - Sync sessions
   - File created
   - File saved
   - File deleted
9. By logging in the profile from a browser, the user can access all the information about deleted files. Investigators can gather information about:
   - Deleted files from trash
   - Author, data of creation and date of deletion of files

10. Track activities on files (ownership, storage location, provenance information).
    Investigators can gather information about:
    - Representative information about files (owner, date of creation, date of edit/modification etc.)
    - Recover information about recent activities

11. Track versioning of files and the option to recover previous versions.
    Investigators can gather information about:
    - Provenance information about file changes, deletion, edition and who has made the changes
    - Recover information about version changes

12. Google Drive RAM analysis: If the client is installed on RAM then the information on RAM can be captured. Investigators can gather information about:
    - User email, Version number, Local sync folder path

13. If the user was logged with a browser session, the info. such as user name and password can be found in RAM
The above mentioned experiment was conducted to examine the digital forensics implementation on cloud environment of Google Drive. The test system was prepared using free hosting plan of 15 GB cloud storage. The Google Drive Client was installed on the local drive and synced with the online cloud storage. The syncing of local drive with cloud storage for the transfer of documents/records to the cloud storage represent the migration of data from client to server in the typical cloud implementation.

The findings from the analysis of the third research question has highlighted the growing threats in the cloud environment and how the digital forensics methods / tools can be applied. The process-approach for handling digital evidence help in identifying challenges and solutions for the different Cloud service models. The experimental representation has presented the implementation of Cloud-based storage ‘Google Drive’ and how the cloud forensics can be applied in the cloud implementation.

4.6. Summary

The analysis and findings presented in this chapter is an outcome of the research work focussing on Digital forensics and Digital archiving. There has not been a lot of research found on the implementation of Digital archiving in Cloud and the complexities of applying Digital forensics to it. The research questions posed in the study has elaborated the promising applications of Cloud technology in Digital archiving as well as the applications of digital forensics vis-à-vis cloud forensics in ensuring the underlying archival parameters of integrity, identity, authenticity and accessibility. The chapter has ended with a presentation of process-based approach to digital evidence handling and the experimental illustration of Cloud based storage (‘Google Drive’).
Chapter 5

Conclusion and future work

This chapter covers the conclusion and the suggestions of the research work. The final discussion in the conclusion would be ended with the contributions of this research work to the academic research as well as the potential for future research on this topic or related areas. Finally, the summary of findings would be presented to highlight the outcome of the research work.

5.1. Conclusion

As the new technologies get extended in the organizations coupled with a growing influence of digital transformation and cloud technologies, the security challenges have also began to appear with an ever increasing new threats and attacks. The detection of emerging threats/vulnerabilities has increased the adoption of new forensic methods and tools to investigate cyber-crimes on one side, and to detect new vulnerabilities on the other side using sophisticated methods, tools and techniques of Digital forensics.

The findings from the research work has presented the step-by-step approach of implementing cloud-based storage for the digital materials. It was also understood from the analysis that the digital forensics methods and tools are extremely important in context of digital archiving. These methods and tools can provide foundation for the implementation of secure, forensically aware and OAIS complaint cloud-based storage. As the cloud technology expansion is growing at an exponential rate, more and more organizations move their operations to the Cloud, the digital forensics implementation can be seen as a key USP for accepting Cloud as a business solution.

The Top 10 threats, attacks and vulnerabilities have been highlighted in the research work to promote the sensitivity of forensics in the cloud-based storage. The process-approach to digital evidence handling can be identified as a core to ensure preservation of digital evidence and protection against data loss or hacks. The emphasis of forensics implementation to cloud-based storage is to ensure identity, authenticity, integrity and accessibility upon which the longevity and sustainability of digital archives is based on.

A process-based approach for handling digital evidence has been presented in the thesis work. The phased process model has highlighted the key challenges and the corresponding solutions for the different service models (PaaS, SaaS, and IaaS) in the
Cloud implementation. It is identified as critical for the organizations to implement robust SLAs between CSP and customers to ensure long-term retention of digital data.

As is the case of digital evidence, it is also important to maintain chain of custody for the digital record. The forensics investigators must strive to preserve the integrity of digital evidence. Any mistakes in the process question the evidence and render it worthless. In the Cloud environment, the authenticity of data can be verified using the Hash algorithm and the integrity of data can be ensured through comparing the digests. If the data is modified accidently or intentionally, the forensic software can detect the modification through comparing the two Hash values. The encryption algorithms such as DES / AES / WPA2 have been found immensely useful for ensuring the confidentiality and authenticity of data during transmission. The forensic software also offers possibility of detecting continued authenticity and non-misuse of data. The proven methods and tools of digital forensics have been found of tremendous use for ensuring authenticity, integrity and accessibility of digital data in the Cloud-based storage. The cloud-storage (e.g. Google Drive) also offers many in-built functionalities to ease digital evidence capture and to ensure the efficient preservation of data.

The final experimentation using Google Drive has highlighted the ease of implementing Digital archiving on Cloud-based storage. The process-approach was applied on the cloud-based storage together with the digital forensics methods / tools to ensure implementation of Digital archiving which was found quite useful and pragmatic.

5.2. Contributions to the research
There has been a very limited amount of research found on the implementation of Digital forensics to the digital archives or e-records hosted in the cloud environment. This thesis work has aimed to investigate the topic in order to develop a broader understanding and to delve deeper into the process-approach for the implementation of digital forensics in Cloud-based digital archives or similar digital repositories. The findings of the research work will be helpful for the custodians of the memory organisations to implement forensically-aware and secure digital repositories/archives.

5.3. Suggestions for future work
The research work in this thesis has only looked at the broader implications of Digital forensics on Cloud specifically in context of hosting Digital archives in Cloud-storage. There is a tremendous potential to research further on the topic such as:
1. Digital forensics framework for implementing cloud-based digital archives: The research on developing a Digital forensics framework for implementing cloud-based digital archive be researched and developed. Currently, there is not framework available for such implementation. There are many different methods and tools but there is no holistic framework for implementation.

2. Audit solution for cloud-based digital archive may be researched and elaborated encompassing the requirements from ISO 27001 Information Security, ISO 22301 Business Continuity and ISO 27043 for Digital forensics.
Bibliography


46. Marangos, N., & Panagiotis, R. M. (2012). Digital forensics in the cloud computing era. *In Globecom Workshops (GC Workshops)*, IEEE.


