

Assessing the Capability of e-Discovery Software Tools

Chirag Vaidya

GradDipCIS, AUT University,
Auckland, New Zealand.

A thesis submitted to the Graduate Faculty of Design and Creative Technologies
AUT University
in partial fulfilment of the
requirements for the Degree of
Master of Forensic Information Technology

School of Computing and Mathematical Sciences

Auckland, New Zealand
2012

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

.....

Chirag Vaidya

Acknowledgements

This thesis has been completed at the Faculty of Design and Creative Technologies of AUT University, New Zealand. I would like to take this opportunity to thank all the people that have provided me with the support that helped me complete this research project. I have received support and motivation from my parents, wife, friends and colleagues. Without your support this thesis might not have been completed.

Firstly, I would like to express my deepest gratitude to my respected supervisor, Dr. Brian Cusack, who has been extremely helpful and offered invaluable support, inspiration and excellent guidance throughout the entire journey of my Master's study. This research would not have been possible without his generous support and I truly appreciate it. I enjoyed our discussions and I feel proud to have had such a supervisor. I would also like to thank the Vound Software team, and especially Peter Mercer who provided me with the limited licence (14 days) for the *Intella 1.5.2* software for this research. I acknowledge the services of Catriona Carruthers and Diana Kassabova who proof-read my thesis.

Finally, I would like to express my sincere gratitude to my dear wife, Divyesha Vaidya, who has inspired and supported me throughout my study. I also would like to thank my daughters, Yesha and Prisha who have supported me during my study. Special thanks to my father Dr. B.M. Vaidya who encouraged and supported me to complete my thesis.

Abstract

Electronic Discovery (e-Discovery) has developed as a process to be managed by investigators and as a practice that has a set of procedures that are peculiar to Electronically Stored Information (ESI). Traditional document management systems have been stable and accessible by manual means. However with the increased use of digital mediums to store information, new techniques have been developed to handle volatile information and its vastly increased quantities. The Electronic Discovery Reference Model (EDRM) is a framework that is widely used as a guideline for e-Discovery processes in investigations. The model provides a systematic guide for actions that start at the information management system and proceed through reproducible steps until an evidential output is achieved. Software tools are also available to perform these investigative steps and to speed the extraction and reporting of evidence. However, the dependability of digital evidence that is collected, analysed and presented in a court using e-Discovery tools has been challenged.

The outputs of e-Discovery processes serve several end-users and are open for scrutiny in a court of law. The main users are those in legal roles who wish to extract reports and presentations from an information management system. Lawyers and other legal advocates require briefs that contain summative information regarding the case at hand. Digital forensic investigators and expert witnesses also require the services of e-Discovery processes and rely on the software tools to deliver full and accurate information that can be substantiated under cross examination. Key issues and problem areas arise from the stability of software, the debates about the reliability of open-source and/or proprietary software, the consistency of different software presentations, and the ability of experts to communicate the use of the software to a court of people unfamiliar with digital processes. Consequently, not only are there many problems surrounding the use of e-Discovery software, but there are also few people who are knowledgeable of both the legal and IT technical requirements of court presentations.

In this research, the most widely used software for e-Discovery processes is reviewed in the literature section and then one of the tools is investigated in the laboratory to assess its characteristics and capabilities. The research question

“What performance can be expected of e-Discovery tools when extracting evidence?” was selected to address the problem of limited knowledge of tool capability. The tool was investigated using each of the phases in the EDRM model, and by testing it in different case scenarios. The results showed the capability of the tool and the scope of such software to assist investigators and others with a legal interest.

To conclude, the overall research conducted confirms that e-Discovery is a legal investigation process that is dependable when the software tools are understood and used correctly. There are many and competing software tools available and each exhibits different strengths and weaknesses. The empirical research study satisfies the aim of testing e-Discovery software to gain greater knowledge of its use. Though pilot testing and case scenarios, the EDRM model was found to be comprehensive and a trustworthy guideline for evidence management. The result of such testing shows a better understanding of a tool’s capability, its effectiveness as a business process, and provides advice for best practices in evidence presentation.

Table of Contents

Declaration	ii
Acknowledgement.....	iii
Abstract	iv
Table of Contents	vi
List of Appendices.....	xi
List of Tables.....	xiii
List of Figures	xv
Abbreviations	xvi

Chapter 1 - Introduction

1.0 Introduction	1
1.1 E-Discovery Issues / Problem Areas	2
1.2 Motivation of the Research	4
1.3 Research Findings	5
1.4 Structure of the Thesis.....	6
1.5 Conclusion.....	7

Chapter 2 - Literature Review

2.0 Introduction	8
2.1 The Importance of E-Discovery	9
2.2 A Definition of E-Discovery	11
2.3 E-Discovery Software	14
2.3.1 EnCase	15
2.3.1.1 Digital Forensics	15
2.3.1.2 Cyber Security.....	15
2.3.1.3 E-Discovery	16
2.3.2 Forensic Tool Kit (<i>FTK</i>).....	16

2.3.3 Vound Software (<i>Intella</i> TM)	17
2.4 Elecronic Discovery Reference Model.....	19
2.4.1 Information Management	20
2.4.2 Identification.....	20
2.4.3 Preservation / Collection.....	20
2.4.4 Processing, Review and Analysis	21
2.4.5 Production.....	21
2.4.6 Presentation.....	21
2.5 E-Discovery Software Capabilities	22
2.6 Expected Software Outputs	24
2.6.1 Business / Client Perspective.....	25
2.6.2 Legal Perspective.....	26
2.6.3 Digital Forensic Investigation Perspective / Technical Perspective	26
2.7 Summary of E-Discovery Issues / Problems.....	27
2.8 Conclusion.....	30

Chapter 3 - Research Methodology

3.0 Introduction	31
3.1 Review of Similar Studies.....	32
3.1.1 Digital Forensics: Validation and Verification in a Dynamic Work Environment	32
3.1.2 e-Discovery Support Tool Design and Implementation of the AGENT Module.....	36
3.1.3 FORZA – Digital Forensics Investigation Framework That Incorporate Legal Issues	37
3.1.4 E-Discovery: Identifying and Mitigating Security Risks during Litigation..	40
3.1.5 Determining Culpability in Investigations of Malicious E-mail Dissemination within the Organisation	42
3.1.6 A Function Oriented Methodology - Searching Function	45
3.2 Research Design.....	46
3.2.1 Evaluation of Similar Studies	46
3.2.2 Review of Problem Areas (from Section 2.7)	47

3.2.3 The Research Question	48
3.2.4 Hypotheses.....	49
3.2.5 Research Plan / Phases.....	50
3.2.6 Data Map	51
3.3 Data Requirements	52
3.3.1 Data Collection	52
3.3.1.1 Search Function Mapping	54
3.3.1.2 Tiered Custodian Data Collection.....	55
3.3.1.3 Tool Test Requirements	55
3.3.1.4 Development of Test Scenarios	55
3.3.1.5 Testing of e-Discovery Tool	55
3.3.2 Data Processing	56
3.3.3 Data Analysis.....	56
3.3.4 Data Visualisation / Presentation.....	58
3.4 Limitations.	58
3.5 Conclusion.....	59

Chapter 4 - Research Findings

4.0 Introduction	60
4.1 Pre-Processing Phase.....	61
4.1.1 EDRM Framework	62
4.1.2 Volume Reduction	62
4.2 Field Findings.....	63
4.2.1 Testing Environments	64
4.2.2 Test case scenario findings	64
4.2.2.1 A – Empirical Low Risk Case Scenario – e-Discovery	65
4.2.2.2 B – Empirical High Risk Case Scenario – e-Discovery.....	71
4.3 Research Analysis	77
4.3.1 Analysis of the Empirical Low and/or High Risk Case Scenario Results	77
4.3.1.1 Test Methodology	77
4.3.1.2 Assessing the Capability of the e-Discovery Tool - Performance Analysis	78
4.3.1.3 Assessing the Capability of the e-Discovery Tool – Search Analysis	80

4.3.2 Assessing the Capability of the e-Discovery Tool	82
4.3.3 Assessing the Capability of an open-source tool for e-Discovery.....	82
4.4 Conclusion.....	83

Chapter 5 - Discussion of Findings

5.0 Introduction	84
5.1 Discussion of the Research Question	84
5.1.1 Main Research Question and Associated Hypotheses.....	85
5.1.2 Secondary Research Questions and Associated Hypotheses.....	86
5.2 Discussion of Findings	96
5.2.1 Discussion of Research Phases Conducted.....	96
5.2.2 Discussion of the Scope of the Selected e-Discovery Tool (Phase 1).....	96
5.2.3 Discussion of the Identification, Preservation and Data Collection (Phase 2).....	97
5.2.4 Discussion of the Test Case Scenarios' Processing, Review and Data Analysis (Phase 3)	98
5.2.5 Discussion of the Data Visualisation / Presentation (Phase 4).....	99
5.3 Discussion of Limitations.....	102
5.4 Discussion of Recommendations : Best Practices.....	102
5.4.1 Focus on Policies for Retention and Deletion	103
5.4.2 Reducing the Cost of Email and Record Management.....	103
5.4.3 Identification of Potential Relevant Information / Data	104
5.4.4 Deploy the Right e-Discovery Tool and a Proactive Approach	105
5.5 Conclusion.....	106

Chapter 6 - Conclusion

6.0 Introduction	108
6.1 Summary of Findings	109
6.2 Answers to the Research Questions	111
6.3 Recommendations for Further Research	111
6.3.1 Testing of Other e-Discovery Tools	112
6.3.2 Testing Different Types of Email Applications and Locations	112

6.3.3 Area of Tools, Future Recommendation	113
6.4 Conclusion.....	114
References	115

List of Appendices

Appendix A.....	123
(e-Discovery Tool – <i>Intella 1.5.2</i> (Vound Software))	123
Appendix 1.....	124
<i>Intella 1.5.2</i> (New Case Processing Options)	124
Appendix 2.....	125
Hardware Configuration (Reference: Table 4.4)	125
(e-Discovery Forensic Workstation01 (core))	125
Appendix 3.....	126
e-Discovery tool Functionality, Feature - Facet.....	126
Appendix 4.....	127
Test Analysis & Findings [Indexing].....	127
Appendix 5.....	129
Configuration Process.....	129
Appendix 6.....	130
e-Discovery Tool Analysis	130
File / Keyword Link Analysis (Example 1)	130
File / Keyword Link Analysis (Example 2)	130
TC01-01 File Structure Analysis:	131
TC01-02 Email Conversation – Timeline :.....	132
TC01-02 File Structure Analysis:	133
TC02-03 Search by Type.....	134
Appendix 7.....	135
e-Discovery tool (<i>Intella 1.5.2</i>) Test Results	135
Appendix 8.....	138
Test # 1 Error while processing zip file	138
Appendix 9.....	139
Test # 2 Unzipping PST File.....	139
Appendix 10.....	140
ESI Checklist: Ref: (www.edrm.net)	140
Appendix 11 (The e-Discovery tool (<i>Intella 1.5.2</i>) Features / Functionality ..	141

Appendix B.....	143
(Open-Source Tool – Digital Forensics Framework (DFF))	143
DFF – Pilot Test Processing & Findings.	144
Test # 1 PST file Processing - Unsuccessful	144
Test # 2 EDRM Dataset File Format Processing - Unsuccessful.....	145
Test # 3 Opening Zip files - Successful	146
Test # 4 PST File Processing - Success	147
Test # 5 EDRM – Data size details from Windows Property	148
Test # 6 File Search – Complexity.....	149

List of Tables

Table 2.1: List of e-Discovery cases	09
Table 2.2: List of e-Discovery cases break down	11
Table 2.3: Characteristics and benefits, <i>EnCase eDiscovery</i>	16
Table 2.4: AccessData Product features	17
Table 2.5: <i>Intella</i> 's Products and their Key Features	18
Table 2.6: e-Discovery Tool Characteristics	23
Table 3.1: Types of Searching Criteria for Keyword-searching	33
Table 3.2: A high-level view of the FORZA framework	39
Table 3.3: Secondary Research Questions	48
Table 3.4: Secondary Research Questions Associated Hypotheses	49
Table 3.5: Collection Source	53
Table 3.6: Data Analysis Key Points	57
Table 4.1: Pre-processing Strategy & Planning	61
Table 4.2: EDRM Framework applied on Test Case Scenarios	62
Table 4.3: Volume Reduction for e-Discovery Process	63
Table 4.4: Case Scenarios	63
Table 4.5: Hardware Configuration	64
Table 4.6: Support Software (e-Discovery tool) Information	64
Table 4.7: AUT-Digital-Forensic-Laboratory Enron-Email-Investigation	66
Table 4.8: Potential source of ESI	66
Table 4.9: Case TC01 brief Special Report	70
Table 4.10: AUT-Digital-Forensic-Laboratory Email-Investigation	72
Table 4.11: ESI Suspects Details	73
Table 4.12: Search Hits for case TC02	73
Table 4.13: Performance analysis	78

Table 4.14: Performance Testing Results on Test Workstation	78
Table 4.15: Performance Testing Result of each Case Scenario.....	79
Table 4.16: Email Analysis including file attachment using Search Methodology	80
Table 4.17: Specific keyword search using Search Methodology	81
Table 5.1: Secondary Research Question 1 and Tested Hypotheses.....	86
Table 5.2: Secondary Research Question 2 and Tested Hypotheses.....	88
Table 5.3: Secondary Research Question 3 and Tested Hypotheses.....	90
Table 5.4: Main Research Question and Tested Hypotheses	93
Table 5.5: Case Scenarios processing, review and analysis.....	98
Table 5.6: List for results exported	99
Table 5.7: List of Potential Source and Basic Information	105
Table 6.1: Summary of findings.....	109

List of Figures

Figure 2.1: EDRM Framework	19
Figure 3.1: Validation & Verification top level model	34
Figure 3.2: Searching breakdown.....	35
Figure 3.3: AGENT total function configuration	36
Figure 3.4: Process flow between the roles in digital forensics investigation	38
Figure 3.5: E-discovery response team's responsibilities	41
Figure 3.6: Network view demonstrating closeness.....	44
Figure 3.7: Compiled from Test Methodology 7, NIST.....	45
Figure 3.8: Research Phases	50
Figure 3.9: Data Map.....	51
Figure 3.10: Collection Guide	53
Figure 3.11: An overview of Search Mapping Function.....	54
Figure 3.12: Analysis Phase Diagram	57
Figure 4.1: The volume reduction of the case TC01 scenario.....	67
Figure 4.2: The visual presentation of file type search results	68
Figure 4.3: The time analysis of target's email communication	69
Figure 4.4: Search hits with Message Hash (MD5)	74
Figure 4.5: Hash Duplicates of case scenario TC02	75
Figure 4.6: Recovered deleted items of case TC02.....	76
Figure 4.7: Time analysis of case TC02.....	76
Figure 4.8: Keyword Search – count in numbers	80
Figure 4.9: Visual evidence for “Thursday night” keyword hits	81
Figure 5.1: e-Discovery analysis process – Data Visualisation	101

List of Abbreviations

AD	AccessData
CEO	Chief Executive Officer
CRM	Customer Relationship Management
CSV	Comma Separated Values
DBX	File extension used by Microsoft Outlook Express
DoS	Denial Of Service
ECA	Early Case Assessment
EDRM	Electronic Discovery Reference Model
EE	Electronic Evidence
EHR	Electronic Health Record
EMC	Electromagnetic Compatibility
EML	Outlook Express Mail Message extension file
EMP	File extension for EMusic File format
ERP	Enterprise Resource Planning
ESI	Electronically Stored Information
FAQ	Frequently Ask Questions
FORZA	FORensics ZAchman framework
FRCP	Federal Rules of Civil Procedure
FTK	Forensic Tool Kit
GB	Gigabyte
IDX	Indexing file used by Microsoft Outlook Express
IDX	Music / Movie Subtitle File
IM	Instant Messenger

IMRM	Information Management Reference Model
ISO	International Organisation for Standardization
IT	Information Technology
NIST	National Institute of Standards and Technology
NSF	Notes Storage Format
OST	Offline Storage Table
PDA	Personal Digital Assistant
PDF	Portable Document Format
PST	Personal Storage Table
RTF	Rich Text Format
SWG	Scientific Working Group
TC	Test Case
USB	Universal Serial Bus

Chapter 1

Introduction

1.0 INTRODUCTION

In relation to digital forensics, Electronic Discovery (e-Discovery or e-discovery or eDiscovery) is the process of finding, collecting and producing Electronically Stored Information (ESI) from computerised systems for litigation. The actual process of e-Discovery is performed by a combination of legal experts and IT experts. The overall e-Discovery process is described in three major steps. In the first step, the legal expert identifies litigation issues and makes a keyword list. In the second step, the IT expert searches for significant information through a systematic process, analyses it and presents the ESI visually in the form of electronic evidence. In the final step, the legal expert reviews and analyses the relevant information and presents it to the court in the form of a report (Lee, Goo, Kim, & Shin, 2011).

The rapid growth of digital information and the fact that it appears as unstructured ESI has increased the importance of e-Discovery investigation. Also, the enormous volume of less formal communication, such as electronic mail (email), wikis, blogs, instant messaging and electronic attachments has increased the workload on those undertaking e-Discovery in corporate environments (Hewlett-Packard, 2011). As a result, the demands in regard to e-Discovery have increased in response to litigation requirements. In December 2006, changes to the Federal Rules of Civil Procedure (FRCP) in the U.S.A., made it clear that ESI is to be preserved as discoverable evidence (Shiekman & Robbins, 2008). In order to comply with the latest regulatory requirements, most businesses, government agencies and even non-commercial organisations have been enforcing a retention policy for legal issues. Likewise, businesses have been warned that altering evidence and/or losing evidence would not be acceptable (FRCP, 2006). Researchers have been debating how to streamline the process of handling e-Discovery queries for litigation hold. Consequently, to facilitate and meet requirements for litigation, organisations have been streamlining their operational efficiency in regard to ESI. Therefore, many commercial e-Discovery tools have

been developed and widely employed by law enforcement organisations and in the private sector to speed up the process of investigation in e-Discovery cases. However, studies that focus on e-Discovery standards and the methodology used to derive evidence using e-Discovery tools are very limited.

In addition, e-Discovery experts are challenged to process large volumes of data while identifying duplicates and backing up large volumes of documents and email. “Electronic messaging or email is the primary target of e-discovery requests” (Hewlett-Packard, 2010, p.8). Therefore, e-Discovery software tools have many key features that assist in processing the unmanageable stream of electronic data and converting it to a manageable size by indexing and removing duplicates (Burgess, n.d.). Hence, the main objective is to assess the capability of e-Discovery tools (a software program that processes and analyses raw electronic materials to locate relevant information) for litigation requirements.

The aim of this chapter is to provide an overview of the research findings and the structure of the thesis. The motivation of the research is briefly discussed in Section 1.2. The main research findings are discussed in Section 1.3. The structure of the thesis is presented in Section 1.4, followed by the conclusion of the chapter.

1.1 E-DISCOVERY ISSUES / PROBLEM AREAS

The survey findings from Fulbright and Jaworski (2009) show that most companies find it difficult to handle e-Discovery issues. According to the survey only

“19% of respondents consider their companies to be well-prepared for e-discovery issues while the vast majority (81%) report being not at all prepared to only somewhat prepared” (Fulbright & Jaworski, 2009, p.14).

There could be many reasons why the vast majority of companies are not prepared for e-Discovery issues. For example, “many law firms either do not have a dedicated e-discovery group or do not have the resources for a particular project” (Volonino & Redpath, 2009, p.274). Various e-Discovery software vendors can provide software and/or services that assist in the e-Discovery process. However,

there are many factors to consider before buying e-Discovery software and/or services.

“Buying e-discovery software is not simple. Courts will not tolerate attorneys with an incompetent e-discovery process that results in missed data or spoliation” (Hall, 2008).

Volonino and Redpath (2009, p. 275) suggest a process for selecting experts and/or consulting companies and to measure e-Discovery software performance and services while negotiating a contract.

In general, the standards for e-Discovery and digital forensics include verification and validation rules, accuracy, auditability, collection methods, protocols and tools (Schuler, 2008). Therefore, The National Institute of Standards and Technology (NIST) in the US established a standard approach for “General Test Methodology for Computer Forensic Tools” via the development of tool specifications, procedures, and testing sets. However, “one of the most heated debates within the field of digital investigations is the lack of standardized methodologies and competencies” (Bayuk, 2010, p.115). Many commercial e-Discovery (investigative) tools advertise their fast, reliable and standardised investigative process as an e-Discovery solution. Likewise, many open-source e-Discovery tools promise a cost-effective process. However,

“as noted by Brian Carrier in 2003, the reliability of an open-source tool must be tested by applying the Dauber guidelines, which focus on testability, error rate, publication for peer review and acceptance by the community” (Bayuk, 2010, p.115).

Baron and Thompson (2007) found that lawyers and their corporate clients face the enormous problem of how to conduct searches for relevant documents in large varied ESI for litigation burdens in an efficient manner. In addition, the conference report from Science and Justice (2010) states that “regulatory trends in forensic science point strongly to the need for exhaustive testing of all findings and tools” (Sommer, 2010, p.12).

Chapter 2, Section 2.7, explores in detail the current e-Discovery problems that have provided the inspiration for this experimental research. In brief, large numbers of tools that support e-Discovery are being released onto the forensic market. However, limited information is provided about the tools approaches for

extracting information from ESI. Hence, the aim of this research is to assess the capability of e-Discovery software using a standard approach.

1.2 MOTIVATION OF THE RESEARCH

The utilisation of high-end technology and ESI has been successful in all business sectors. Therefore, it is important for businesses to maintain a record management policy for their day-to-day operation. It is also equally important to retain all this information for litigation purposes. Hence, investigators from both the private and public sectors are relying heavily on e-Discovery tools to gather, assess and analyse ESI as part of digital evidence.

“The e-Discovery market has been fragmented, lacking a fully integrated solution, and instead relying on multiple point solutions, which breeds inefficiencies, causes delays and increases risks, and ultimately, costs” (Pasadena, 2010).

The complexity of digital forensics, the e-Discovery process and forensics tools means that standardised frameworks for digital investigation need to be devised (Beckett & Slay, 2007). The capabilities of and the results provided by e-Discovery software form the basis of the legal hold. This is the motivation for assessing the capabilities of e-Discovery tools by testing and measuring performance and outcomes. Additional motivation for this research is the following finding:

“A lack of good e-discovery capabilities, as well as inadequate technologies that support e-discovery, can drive up labour, legal and other costs. For example, some discovery systems produce up to 1,000 times more content than is actually required, increasing discovery costs unnecessarily” (Osterman Research, 2011, p.1).

If the e-Discovery tool produces irrelevant content and/or information, this would result in higher case investigation costs. Osterman Research (2011) discussed a few e-Discovery cases and decisions that are relevant to consider in the context of e-Discovery. In general, e-Discovery software is designed specifically for litigation support, as it searches and processes ESI and presents acceptable evidence for litigation. At this point, the challenge is to assess the performances of

e-Discovery software. Therefore, this study is aimed at testing the performance of the available e-Discovery tools by applying an appropriate methodology.

Many businesses already have realised that a pro-active approach will speed up the process of e-Discovery for litigation requirements and will minimise the time and cost of an investigation. The potential advantages of pro-active e-Discovery management procedure are believed to improve business operation. In addition, testing the capabilities of an e-Discovery tool will increase confidence in selecting the right tool and will save time and effort. These are all motivations behind assessing the capability of the available e-Discovery software in this research.

1.3 RESEARCH FINDINGS

This research introduces the Electronic Discovery Reference Model (EDRM) guidelines as a benchmark framework for this investigation (Section 2.4). In addition, the features and functionality of e-Discovery tools are evaluated theoretically, as the individual tool capabilities are discussed in Section 2.5. The research findings of this project are the result of empirical work that has been carried out based on the methodology discussed in Chapter 3. An empirical research method was used for this project. The research findings reported in Chapter 4 are based on the research phases discussed and illustrated in Chapter 3. These phases are as follows: Phase One: Assess the scope of the selected tool, Phase Two: Test the performance of the e-Discovery tool. Phase Three: Assess the capability of the selected e-Discovery tool, and Phase Four: Presentation (see Figure 3.8). The overall e-Discovery investigation is based on a pre-processing strategy, planning and two major case scenarios that demonstrate the selected e-Discovery tools capabilities, performance, searching abilities and its ability to produce reports for a litigation hold. The testing workstation is prepared and configured for pilot testing, while the case scenarios, using various methods according to test specifications, are reviewed in Chapter 3.

Chapter 4 presents and discusses the empirical research findings of this project. The e-Discovery strategy based on the Electronic Discovery Reference Model (EDRM) is presented in Section 4.1. The raw results of the research findings come from the case scenarios, performance testing, analysis and

verification of the e-Discovery tool (See Section 4.2). The experiential research findings for assessing the capability of the e-Discovery tool (*Intella 1.5.2*) are reported (Section 4.3).

1.4 STRUCTURE OF THE THESIS

The structure of the thesis consists of six major parts. Chapter 1 introduces the motivational factors for the e-Discovery investigation and highlights the research areas including current e-Discovery issues.

Chapter 2 presents a literature review of relevant research, critical evaluation of similar studies and the findings of other academic research papers. The chapter reviews the definition and importance of the e-Discovery process and the characteristics of e-Discovery tools. It also identifies the importance of the EDRM stages of e-Discovery in relation to the digital forensic investigation process. A brief explanation of each EDRM stage is presented (Section 2.4) in order to understand the importance of the framework for the e-Discovery forensic process that has been accepted by many large organisations. In addition, Section 2.3 describes the key features, functionality and benefits of leading software in the industry (*EnCase*, *FTK* and *Intella*). The expected output from e-Discovery tools is discussed from a few different perspectives in Section 2.6. Finally, the chapter provides a summary of the e-Discovery issues identified (Section 2.7).

Chapter 3 identifies and builds the foundation of this empirical research by reviewing a few similar studies (Section 3.1) and studies their methodologies. The chapter then sets out the methodology for this study and the main research question in Section 3.2.3. The data requirements and the limitations of the research are also reviewed and discussed in Section 3.3 and Section 3.4 respectively.

Chapter 4 reports the research findings. The chapter is split into four major sections that present the strategy and planning (Section 4.1), the findings from the experimental work (Section 4.2), the research analysis (Section 4.3) and the final section, 4.4, provides a conclusion. The main findings are summarised and presented graphically in this chapter.

Chapter 5 uses the research findings presented in Chapter 4 to assess the capabilities and performance of the selected e-Discovery tool (Section 5.2). This

chapter discuss the evidence obtained from the tested scenarios and related to the main research question and associated hypotheses defined in Chapter 3. Overall, Chapter 5 provides an in-depth discussion of those findings and refers back to previous relevant sections and to the recommendations for best practices (See Section 5.4 for more details).

The final chapter of this thesis summarises the research findings, answers the research questions (Section 6.2) and discusses the possible directions for future research (Section 6.3). The last section is the conclusion of the thesis.

1.5 CONCLUSION

Chapter 1 presents the context and the main motivation for the research. This chapter focuses on the current e-Discovery issues and the purpose of assessing the capabilities of e-Discovery tools.

This research has focused much of its efforts on finding out about e-Discovery standards, benchmarks, methodology and the foundations of the e-Discovery process. The motivation behind the research is also explained in this chapter. The research findings of the research are clarified, including e-Discovery procedures and performance of e-Discovery tools.

Chapter 2

Literature Review

2.0 INTRODUCTION

This chapter introduces the concept of electronic discovery (e-Discovery). The key features and benefits of industry-leading, e-Discovery software tools such as Guidance Software (*EnCase*), Forensic Toolkit (*FTK*) and Vound Software (*Intella*) are identified so that the scope of e-Discovery software is defined. The importance of the EDRM for the analysis of e-Discovery processes and subsequent sub-processes is also discussed.

e-Discovery is a business process that mines the reports of Digital Forensic investigations and presents them ready for courtroom use. “The ability to request and/or produce electronic evidence can mean the difference between winning and losing your next case” (Arkfeld, 2006, p.2).

“Electronic discovery (also called e-discovery or e-Discovery) refers to any process in which electronic data is sought, located, secured, and searched with the intent of using it as evidence in a civil or criminal legal case” (Pateriya, Mishra, & Samaddar, 2011, p.504).

Such data can be found for example, in electronic text, images, databases, electronic mail (email), all email attachments, audio/video files, voice mail, cell phone messages, digital photographs and backups. FRCP, Rules 16, 26, 33, 34, 37 and 45 provides a complete set of e-Discovery amendments, with the accompanying Advisory Committee notes. A few examples of e-Discovery issues are investigation of confidential information leaks through e-mail, employees’ misuse of computers, electronic harassment and/or any unauthorised activity that breaches company policies and procedures (Netsecurity Corporation, 2008).

The aim of this chapter is to define terms and to familiarise the reader with the architectures and concepts used in the field of digital forensic e-Discovery. Section 2.1 discusses the importance of e-Discovery, in section 2.2, a few definitions of e-Discovery as provided by various literature sources are discussed. Section 2.3 reviews the key features of e-Discovery tools. Section 2.4 provides an

overview of the EDRM and a brief explanation that helps to identify the importance of the e-Discovery process guidelines and framework for obtaining effective results in an organised manner. Sections 2.5 and 2.6 review e-Discovery software capabilities and identifies the key features and expected software outputs from a few different perspectives. Finally, section 2.7 summarises the key problems and issues raised in the reviewed literature.

2.1 THE IMPORTANCE OF E-DISCOVERY

Today, e-Discovery is in effect a mandatory requirement in the United States. “This is because of a series of laws which require companies to be able to produce documents, media, and communications for oversight activities” (Forte & Power, 2006, p.8). In order to comply with the latest regulatory requirements, e-Discovery demands from courts and stakeholders are increasing rapidly.

“Industry organisations have begun to make some initial attempts at creating standards and best practices. The Sedona Conference has a number of guidelines and best practices recommendations available for e-Discovery topics, including search protocol and choosing an e-Discovery vendor” (Knox & Dawson, n.d.).

In answer to these demands, the EDRM Project was designed to help organisations manage the process of e-Discovery from the initial stages of ESI searching through to evidence presentation (EDRM, n.d.). According to *ISO 9001: A Foundation for E-Discovery*, in order to maintain certification, an organisation must implement management responsibilities, internal quality audits, monitoring and measuring, continual improvement, corrective and preventive actions (Knox & Dawson, n.d.).

Table 2.1: List of e-Discovery cases (Compiled from Civil Discovery & Privilege Law, n.d.)

	Case	Code / When	Brief Description
	<i>Weatherford U.S., LP v. Innis</i>	<i>June 2, 2011</i>	<i>Protocol for computer exam.</i>

<i>National Day Laborer Organizing Network v. United States Immigration and Customs Enforcement Agency</i>	<i>(Southern District of New York (SDNY). 2011)</i>	<i>Usable ESI must be searchable and include metadata. Specific metadata identified.</i>
<i>DeGeer v. Gillis</i>	<i>(December 8, 2010)</i>	<i>ESI Cost shifting factors.</i>
<i>Wilson v. Thorn Energy</i>	<i>LLC, 2010 WL 1712236 (SDNY March 15, 2010)</i>	<i>Evidence exclusion sanction for failure to preserve data on flash drive by backup copy. "Safe Harbor" provision does not apply.</i>
<i>Accessdata Corp. v. ALSTE Tech. GMBH</i>	<i>2010 WL 3184777 (D. Utah January 21, 2010)</i>	<i>Form of production cannot be less useful than the form which is maintained by producing party.</i>
<i>Zubulake v. UBS Warburg</i>	<i>SDNY 7/20/04</i>	<i>UBS failed to preserve relevant e-mails, even after receiving adequate warnings from counsel, resulting in the production of some relevant e-mails almost two years after they were initially requested, and resulting in the</i>

			<i>complete destruction of others. For that reason, Zubulake's motion is granted and sanctions are warranted.</i>
--	--	--	---

Table 2.1 shows the list of a few e-Discovery cases. The brief description of each case describes the reason for and the importance of e-Discovery. In addition, Kroll Ontrack (2007) discusses judicial opinion issued in 2007 and the importance of new FRCP that focuses on what to do when parties fail to play by the new rules. According to Kroll Ontrack, (2007), approximately 105 e-Discovery opinions had been reported since December 1, 2006.

Table 2.2: List of e-Discovery cases break down (Compiled from www.findlaw.com)

E-Discovery Cases (major issues)
• 25% of cases addressed discovery requests and motions to compel
• 24% of cases addressed spoliation/sanction
• 23% of cases addressed issues involving the form of production
• 9% of cases addressed preservation/litigation holds
• 7% of cases addressed attorney-client privilege and waiver
• 6% of cases addressed production fees
• 6% of cases addressed admissibility of electronic evidence

Table 2.2 summarises the major issues involved in these cases. Therefore, understanding the impact of the rules for e-Discovery and their use in litigation is very important to many organisations.

2.2 A DEFINITION OF E-DISCOVERY

Matthews defines e-Discovery as “simply the process of locating, collecting and organising relevant electronically stored information, usually for litigation” (Matthews, 2010). At this point, the controversial argument is that e-Discovery is not a simple process, it requires many software tools, techniques, and experts to

collect information. It might be very complicated if the strategy of e-Discovery is not applied in an organised manner. Therefore many organisations agreed that the EDRM could be used as a framework for the systematic investigation of ESI. In general, e-Discovery is the investigative process of collecting information from computers, networks, software, hardware, computer peripherals and every single item that contains electronic data.

The Gartner research document states that “discovery is a form of legal interrogatory –a way in which opposing parties in a lawsuit can elicit admissible evidence” (Gartner, 2007). In addition, it mentioned the definition from Black’s Law Dictionary, for discovery being,

“the act of, or process of, finding or learning something that was previously unknown. It is the compulsory disclosure, at a party’s request, of information that relates to the litigation” (Gartner, 2007, p.2).

“During routine legal discovery, a plaintiff is entitled by law to have access to documents stored in corporate memory” (Barker, Cobb, & Karcher, 2008, p.181). Here, the term “corporate memory” refers to a collection of all ESI stored on different devices that include mobile phones, Personal Digital Assistants (PDAs), voice mail, Electronic Mail (email), Instant Messages (IMs), text messages, digital photos and all electronic documents attached to emails including video clips (Barker et al., 2008).

Osterman Research (2010) explains the difference between “Discovery” and “E-Discovery”. It provides key information about e-Discovery and its priority to businesses. According to Osterman Research (2010),

“Discovery is the compulsory disclosure of pertinent facts or documents to the opposing party in civil action, usually before a trial begins and e-discovery is simply the extension of this well-established process to the electronic content that an organisation might possess, including email messages, instant messages, word processing files, spreadsheets, presentations, purchase orders, contracts, social networking content, files stored in collaboration systems, and all other electronic content to which an organisation might have access” (Osterman Research, 2010, p.2).

Therefore, e-Discovery investigation requires digital forensic knowledge and experience to collect and organise important files, formats and verification of the ESI. These may include security domains for authentication, recovery of hidden ESI, lost and/or damaged evidence from suspect machines and/or from a computer network. From a technical perspective, e-Discovery engages with specific investigations of all types of electronic documents that require identification, location, preservation and retrieval for law enforcement. The pro-active e-Discovery approach is to support policies, procedures and corporate compliance to minimise risks and costs. Furthermore, the e-Discovery investigation may extend to specific computer and/or network locations, on devices such as e-mail servers, file servers, iPods, desktops and hard drives. Similarly, dedicated software office applications directly store information in different electronic formats, for example, Microsoft Outlook stores all information in a Personal Storage Table (.PST) file. There are many technical and legal requirements related to the e-Discovery process. The “*Buyer’s Guide for IT Professionals*” explains:

“E-discovery is a process that involves many stages and collaboration with internal and external resources, including inside legal counsel, IT, and outside counsel who must work together for success” (Guidance Software Inc., 2010, p.2).

The guide provides many technical aspects in its “Product Evolution Checklist” that should help an Information Technology (IT) Professional choose appropriate solutions for their own e-Discovery process. It describes a 10-step e-Discovery process and its requirements that should reduce the burden for the IT departments in organisations. Overall, this guide claims that their product’s strengths compare with other products for in-house discovery software.

The aim is to minimise the scope of an e-Discovery investigation and thereby reduce the cost of electronic discovery of evidence to be presented for the legal process. For example, companies should consider implementing a document retention policy for record management. The Chief Justice Guidelines define ESI as “any information created, stored, best utilised with computer technology of any type” (Losey, 2009, p.64). Losey (2009) describes a few different methods and types of equipment for storing electronic information. However, lawyers and forensic investigators need to follow many guidelines whenever e-Discovery is involved in a case. For example, the Chief Justice Guideline Three describes eight

different categories of information that “judges may want to order the parties to provide to each other so as to facilitate agreement on e-Discovery issues” (Losey, 2009, p.65).

A definition of e-Discovery contains few words but it still encompasses many processes and sub-processes used to collect information for litigation purposes, using a strategic plan and procedures that are legally acceptable. In this chapter, e-Discovery is defined as a set of techniques used to perform the e-Discovery process and present acceptable evidence for law enforcement.

2.3 E-DISCOVERY SOFTWARE

According to Markoff (2011), “e-Discovery software can analyse documents in a fraction of the time for a fraction of the cost”. At this point, the aim of using e-Discovery software is to minimise cost, time and effort in producing relevant information as output. e-Discovery software provides a better and speedy process for producing results when compared to traditional manual discovery. The traditional approach for discovery consumes large amounts of time, money and manpower to process the evidence for law enforcement. “Smarter than you think”, a news article in “The New York Times”, claimed that

“the studios examined six million documents at cost of more than \$2.2 million, much of it to pay for a platoon of lawyers and paralegals who worked for months at high hourly rates”
(Markoff, 2011).

e-Discovery software is designed specifically for litigation support. It searches and processes ESI and presents a report as evidence for law enforcement. There are many different industry-leading software products that will do e-Discovery on demand and/or as it is required. The main purpose of e-Discovery software is to facilitate a legal hold, pre-collection analysis, collection, preservation, processing and reporting of relevant information. Another benefit is to reduce costs and risks, and to improve the chances of success of any information technology project. The aim of this chapter is to identify the key components of e-Discovery software, and to assess the capabilities of selected e-Discovery tools by testing their performance and functionality.

To minimise cost, time and effort and to speed up the legal process, litigation panels and forensic experts are required to use the best available e-Discovery software tools. The right e-Discovery software tool has many key features and supports legal teams in Early Case Assessment (ECA). Due to the constantly increasing volume of digital data, it is becoming harder and harder to find and/or collect relevant information for a legal process and this is a challenge for legal teams.

Subsections 2.3.1 and 2.3.2 discuss some industry-leading computer forensics tools (*EnCase*, *FTK*) that provide separate tools for e-Discovery service and support. The range of e-Discovery software has created many opportunities for forensic investigation and law enforcement, although only few software applications are widely popular and accepted by court. The remainder of this section briefly discusses the key features of the most popular forensic suites.

2.3.1 EnCase

EnCase is a forensics suite of tools created by Guidance Software and certified by the National Institute of Standards and Technology (NIST). It is widely used worldwide by law enforcement agencies and private computer forensic examiners. There are many software products and services available on its official site; some of the key tools are widely used in digital forensics and cyber security and for e-Discovery.

2.3.1.1 Digital Forensics

This is a single tool capable of conducting large-scale and complex investigation processes using in-depth analysis and forensically sound acquisitions. It has advanced productivity features that enable it to find information, create cases, transfer evidence files and generate reports for law enforcement. This software is mainly used by commercial firms, government agencies, private investigation groups and law enforcement to conduct thorough investigations for digital forensics.

2.3.1.2 Cyber Security

The Cyber Security tool is designed to detect malicious code and to proactively respond to network threats across global networks. Significant advanced key features are added that allow for the identification of new threats. For example,

some advanced algorithms (for measuring the entropy, code analysis and memory analysis) are designed to identify new attacks and to provide solutions similar to those provided by anti-virus software (Guidance Software Inc., 2012).

2.3.1.3 E-Discovery

The e-Discovery tool is designed to “...provide an integrated, in-house e-discovery solution that is both forensically-sound and court-validated, minimising risk and reducing cost by up to 90%” (Guidance Software, n.d.). Distinct features of the e-Discovery software are the proactive assessment approach for analysis and the generation of a first-pass review to improve case strategy (Guidance Software, n.d.). For example, the *EnCase eDiscovery* tool can identify data custodians, craft efficient search terms and conditions, investigate electronic mail, and extract, process and export to standard attorney review platforms.

Table 2.3: Characteristics and benefits, EnCase® eDiscovery (Source: www.guidancesoftware.com)

Key Benefits	Key Features
<ul style="list-style-type: none"> ✓ Reduce Risk ✓ Cost Reduction ✓ Efficient Business Process 	<ul style="list-style-type: none"> • Single unified solution • Legal hold • First-Pass review analysis • Pre-collection analytics • Collection and preservation • Judicially accepted & defensible • Readily scalable • Licence option with pay-per-use

2.3.2 Forensic Tool Kit (FTK)

Forensic Tool Kit (*FTK*) is a forensics suite of tools created by AccessData Corporation and is used by law enforcement agencies and private computer forensics examiners worldwide (AccessData, n.d.). *AccessData (AD) eDiscovery* is a leading e-Discovery software tool for the search, identification, collection, preservation and processing of ESI residing on desktops, laptops, email servers and removable storage media. *AD eDiscovery* maintains the integrity of the collected data, including file metadata. Additionally, its extensive processing

capabilities enable post-collection culling and de-duplication of data. The features of *AD eDiscovery* software are designed for many categories such as corporate e-discovery, law firm e-discovery and service provider e-discovery (AccessData, n.d.). Table 2.4 presents the key features for three products from AccessData.

Table 2.4: AccessData Product Features (Source: www.accessdata.com)

Product	LIT.Hold	In Place Search	Forensic Collection	Processing	Internal Review	Load File TIFF conversion	Final Review	Trial Support
AD eDiscovery	✓	✓	✓	✓	✓	✓		
AD ECA				✓	✓	✓		
AD Summation							✓	✓

Recently, AccessData Group introduced *AD Summation Case Vantage* with enhanced features. According to Lee (2011),

“Case Vantage is a database-driven, full-featured legal review software platform that provides secure internet access to case data and review tools using nothing more than a standard web browser. It has an intuitive web-based interface, and is the ideal legal review platform for working with distributed teams and outside parties” (Lee, 2011, p.2).

2.3.3 Vound Software (*Intella*TM)

Intella created in 2008 by Vound Software is a “...powerful indexing search engine software with visualisation features for search and review email and electronically stored information (ESI) to collect important evidence” (Forensic Computer, n.d.).

“It is ideally suited for use by enterprise, law enforcement, regulatory agencies, and law firms in civil, criminal, or policy related investigation” (Vound Software Inc., 2011, p.5).

The software demo on their website illustrates how quickly and easily the tool can search ESI to find critical evidence. It also provides visual analytic features. Due to its innovative email investigation features and functionality, *Intella* e-Discovery software is quickly becoming the preferred tool for enterprises, professional service firms, law enforcement and government agencies. Vound Software has

introduced many products such as *Intella Desktop* and updated versions with different names. However, the most recent product is *Intella TEAM*.

Table 2.5: *Intella* Products and their Key Features (Compiled from Vound Software Inc., 2011)

Products	Function / Key Features
<i>Intella</i> TM Desktop	Index and search multiple email, file types and metadata
<i>Intella</i> 10 Desktop	Limited indexing to 10GB at a time. (for smaller cases)
<i>Intella</i> Viewer	An optional companion for the above two products.
<i>Intella</i> TEAM	Enables multiple individuals (reviewers, investigators, paralegals)
<i>Intella</i> TEAM has the following two components that perform critical functions:	
<i>Intella</i> TEAM Manager	<ul style="list-style-type: none"> • Indexing & preparation of the case data or evidence • Sharing of the case data among others • Combining the work product of others
<i>Intella</i> TEAM Reviewers	<ul style="list-style-type: none"> • Independent search, filter, bookmark, tag and comment

Table 2.5 shows *Intella* product features for e-Discovery investigation with their critical functions. The key features are indexing for email and hard drives, keyword searching for specific items in emails with comprehensive visual cluster map that can be presented in different report formats such as PDF, RTF and CSV files. In comparison with the other two well-known industry tools (*EnCase* & *FTK*) many of its basic key features and benefits are found to be very similar. For example, it provides legal hold, pre-collection analysis, internal review, processing and searching facilities.

An alternative option is to research an open-source e-Discovery tool for assessing similar capabilities. The philosophy behind using an open-source approach is that anyone can undertake e-discovery without buying a licence and can get the same output with minimum cost. However, identifying open-source tools is difficult, time-consuming and a complex process in comparison to using commercial tools.

2.4 ELECTRONIC DISCOVERY REFERENCE MODEL

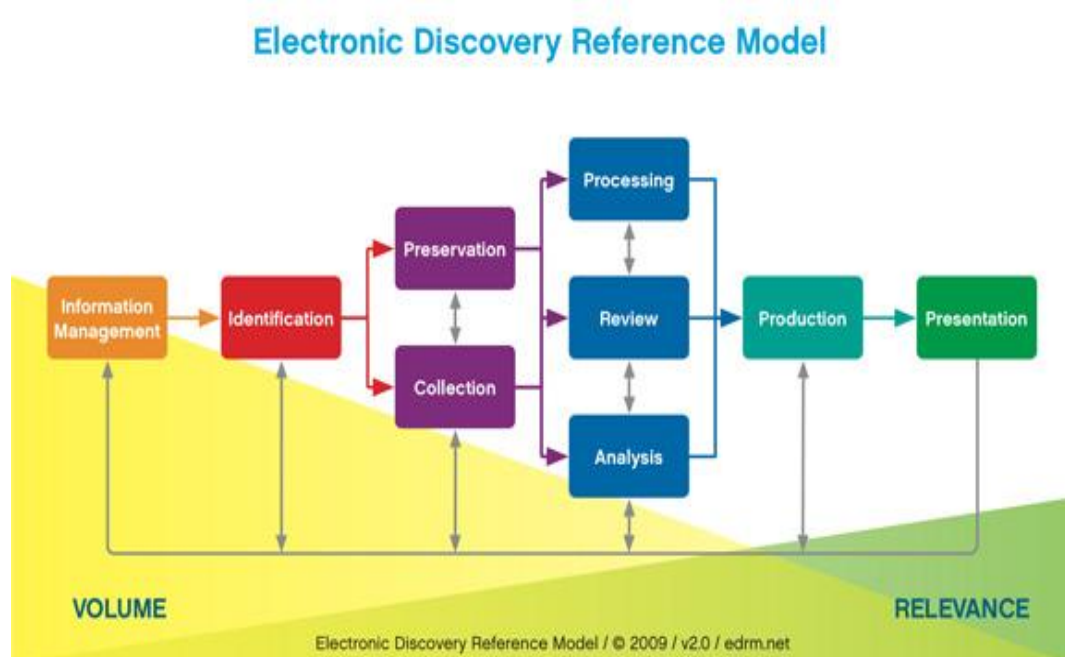


Figure 2.1: EDRM Framework (Source: www.edrm.net)

The Electronic Discovery Reference Model (EDRM) Project is designed to help organisations manage the process of e-Discovery from the initial stages of searching and collecting ESI through to its presentation. The team that developed the EDRM was facilitated by George Socha and Tom Gelbmann and included 62 organisations, software developers, law firms, professional organisations and large corporations. The entire reference model is made up of interconnected detailed sub-processes or sub-frameworks. According to the official EDRM website;

“EDRM develops guidelines, sets standards and delivers resources to help e-discovery consumers and providers improve quality and reduce costs associated with e-discovery” (EDRM, n.d.).

In brief, EDRM is a project that has been created by multiple organisations to establish set protocols, guidelines, recommendations and procedures in order to assist individuals involved with an e-Discovery request. A brief description of all its stages and their significance to the e-Discovery process are provided below.

2.4.1 Information Management

The EDRM introduced a new sub-framework known as the Information Management Reference Model (IMRM). Its draft guide provides information about a “...practical, flexible framework to help organisations develop and implement effective and actionable information management programs” (EDRM, n.d.). The beneficiaries of this project guidelines are legal advisors, business leaders and stakeholders, information technology users and digital forensic investigators. The process that the e-Discovery project deals with involves day-to-day management of data from different electronic sources for preservation. It includes daily backup of data by using modern backup software, indexing files structure, recovering data from deleted sources and many other activities that can be produced in the form of reports for court room use.

2.4.2 Identification

A digital forensic investigation locates relevant information related to a particular case. It is required to “develop identification and strategy plan” to identify “potential source of information” described as ESI and considered relevant.

2.4.3 Preservation / Collection

An important step in the digital forensic investigation and e-Discovery process is the preservation of information and the collecting of evidence for legal purposes. During this stage, digital forensic tools and techniques are used to ensure that all collected ESI is not altered and/or destroyed, either intentionally or unintentionally. “Cases have literally been won or lost based on the practices of preservation and recovery followed by organisations” (Matthews, 2010). Therefore, it is the duty of the investigator to preserve relevant information using preservation strategies and to stick to the required policies. For example, in their daily operations, many companies destroy and/or delete unnecessary files to save electronic resources. However, “ESI destroyed under an organisation’s document destruction policy prior to, or reasonably anticipated litigation is exempted under the FRCP 37(e) Safe Harbor Provision” (Chisholm, 2010).

2.4.4 Processing, Review and Analysis

These are the most complicated and time-consuming stages, where all information that is relevant to the case is examined thoroughly. All collected and preserved digital information relevant to the incident is fully verified and examined at this stage. For example, the investigator may need to search the content of all devices associated with the case using different software tools and techniques such as keywords searching, computing hash values, performing an email search, verifying file signatures, identifying code pages, searching for internet history files and performing a comprehensive search. As per the EDRM “Processing Guide”, the processing stage can be further “broken down into four main sub-processes, namely: Assessment, Preparation, Selection and Output” (EDRM, n.d.). The review and analysis stages are also further divided into sub-categories to facilitate further findings and research.

2.4.5 Production

Production is an essential stage of e-Discovery where important and relevant information is prepared and produced in a format that is appropriate for law enforcement and court room presentation use. For example, all digital information such as images, office suite documents including databases, email, and mobile phone record details need to be presented in forms and reports (Chisholm, 2010). The purpose of production is to

“...prepare and produce ESI in an efficient and usable format in order to reduce cost, risk and errors and be in compliance with agreed production specifications and timeliness” (EDRM, n.d.).

It includes tracking documents such as emails, downloads, login details, indexing, sharing, audit log information and many more searches depending on case requirements.

2.4.6 Presentation

Presentation is an important step in the e-Discovery and digital forensic investigation process. The aim of presentation is to provide, “native and near-native” forms of relevant information and/or ESI to audiences, jury and court. “Simply stated, if the material is not displayed properly in front of the jury, then all the effort is for naught” (EDRM, n.d.). According to the EDRM reference

guide, this step deals with trial and deposition, strategy plan, legal analysis, client and witness interview and meetings, court room process, preparing exhibits, testing and delivery, presentation, acquisition and storage.

According to Adam (2011), the EDRM is the foundation of the e-Discovery process and it enables investigators to perform successful e-Discovery. Successful e-Discovery requires a strong foundation in order to track ESI in an organised manner. Hence, EDRM, and all its stages mentioned above, help to maintain the strategy and procedures for e-Discovery. On the other hand, there is a large volume of ESI that needs to be stored in archives, particularly e-mails, attachments and mobile communications, for future litigation purposes. To minimise expense and inefficiencies when dealing with legal inquiries and to plan for e-Discovery process, the EDRM will play a significant role.

“The EDRM Data Set Project provides industry-standard, reference data sets of electronically stored information (ESI) and software files that can be used to test various aspects of e-discovery software and services” (EDRM, n.d.).

For example, *EDRM Enron Email Data Set v2* supports multiple email formats.

To sum up, the EDRM Guide focuses on many aspects of the e-discovery process such as search, retrieval and production of ESI described in the EDRM Model. Figure 2.1 is intended as a basis to discuss the important stages and to analyse the EDRM framework and its significant use in the e-Discovery process. However, there might be many other different ways of analysing the e-Discovery process that would lead to different solutions.

2.5 E-DISCOVERY SOFTWARE CAPABILITIES

According to NIST, “there is a critical need in the law enforcement community to ensure the reliability of computer forensic tools” (NIST, 2011). The reliability of computer forensic tools can be assessed through their capabilities. In general, e-Discovery software is designed to process emails, to analyse metadata, archive files, backup devices and any form of electronic documents. In other words, the software is capable of supporting the e-Discovery process. Table 2.6 presents the key features of some industry-leading, e-Discovery tools.

Table 2.6: e-Discovery Tool Characteristics (Compiled from Vound Software, Guidance Software & AccessData)

e-Discovery Tools	Key Features / Characteristics
Vound Software (<i>IntellaTM</i>):	<ul style="list-style-type: none"> • Preview email and data files for investigation and e-Discovery • Gain deeper insight through visualization • Search email, attachments, embedded images, archives, headers and metadata • Drill deeply using <i>IntellaTM</i>'s unique facets • Group and trace email conversations • Export results in a choice of formats for later use or reporting
<i>EnCase eDiscovery</i> :	<ul style="list-style-type: none"> • Extensive searching capabilities • Unique advanced search term analysis • Screening & search reports • Linear review with hit highlighting • Email threading & conversation viewing • User-defined tagging, comments and classification • Full audit trail for defensibility with security access
<i>AccessData –FTK</i> :	<ul style="list-style-type: none"> • Pre-collection auditing on-site or remotely • Advanced keyword searching and filtering • Creates native Personal Storage Table (PST)s and Notes Storage Format (NSF)s from email servers • Custodians create data and smart restart functionality • Backup tape extraction and media dialog collecting ESI • Data collection from workstations, laptops, network shares, email servers, database and

	30+ structured data repositories, including SharePoint and Enterprise Vault.
--	--

It can be seen from Table 2.6 that the individual e-Discovery forensic suites have their own unique functionality. In addition, *EnCase eDiscovery version 4*, provides a “Web Interface and Early Case Assessment” feature that is very useful for both legal and IT Professionals. For example, early case assessment is possible through an internet connection for searching keywords, viewing and analysing data. “In contrast to other products, in which analysis and review can be performed only after collection and processing is complete” (Pasadena, 2010). The capabilities and the results obtained by using e-Discovery software form the basis of a legal hold, showing the importance of analysing the capabilities of e-Discovery software. A case study carried out by a top communication company found that e-Discovery when performed with the help of *AccessData* was much faster than when performed with *EnCase*.

“Their evaluations against *EnCase eDiscovery* and the various Clearwell modules confirmed for this company’s Discovery team that *AD eDiscovery* delivers an innovative and cost-effective framework that enables the speed, accuracy, efficient workflow and forensic integrity they were looking for” (AccessData Group, 2011, p.5).

The information presented in Table 2.6 provides the motivation for assessing the capabilities of at least one of the selected e-Discovery tools by testing in order to make appropriate assumptions and to achieve the best possible outcomes in the e-Discovery process. The next two sections discuss the expected outputs from the e-Discovery software and also some issues related to the e-Discovery process.

2.6 EXPECTED SOFTWARE OUTPUTS

With an increasing number of corporate investigations, regulatory audits and cases of internal fraud, the e-Discovery process can be seen to improve business processes. A correct internal e-Discovery setup could improve business processes

and reduce risk (Guidance Software, n.d.). The aim of this chapter is to identify the key features of e-Discovery software tools, the importance of the EDRM framework and to assess the capabilities of e-Discovery software in terms of complexity (data processing difficulty), functionality and reliability.

Any cyber fraud needs proper investigation that involves enormous cost for hiring special agencies and/or forensic experts to solve the issue and collect the evidence for law enforcement within a specific period of time. Therefore, information obtained by assessing software tools capabilities, will reduce the time and effort for selecting correct options and techniques for a specific investigation. For example, tool X is better in an email search investigation when compared to tool Y. Expectations of e-Discovery software outputs may benefit business, legal parties, and technical professional.

2.6.1 Business / Client Perspective

Business owners expect that an e-Discovery software should be cost-effective, simple and well organised so that future e-Discovery queries can be easily maintained for litigation. However, the expected outcome from e-Discovery software would be more effective if factors such as affordability, versatility, search accuracy and easy deployment are present (“5 Tips for Choosing e-Discovery Software,” 2010). According to Guidance Software,

“*EnCase eDiscovery* provides the most efficient and cost effective processing capabilities that greatly reduce the data set collected eliminating non-responsive ESI based on advanced search techniques. The end result delivers targeted and responsive results” (Guidance Software, n.d.).

The other two software applications also claim similar cost-effective features and capabilities. For example, *EnCase eDiscovery* claims that their software “speed is 50 times faster than other solutions” (Guidance Software, n.d.). Likewise, many companies choose Vound Software (*Intella*) for its powerful indexing with visualisation features for searching and reviewing email. These types of features and benefits are appealing from the business point of view. However, to prove the speed and effectiveness of software requires testing of tools and comparison of capabilities.

2.6.2 Legal Perspective

The successful implementation of retention policies in the organisation and the e-Discovery software solution are the major source for legal defensibility and evidence preservation. Gartner advises organisations to consider certain key features in offering e-Discovery tool such as email classification, user access to archived email, legal and information discovery capabilities (Favro, 2012). The *Intella* tool by Vound Software, claims that its capabilities are user friendly due to the visual analytic feature and the ease of performing searches of emails and ESI.

From a non-technical perspective, e-Discovery software should provide user-friendly features, and should easily provide a locus of attention for non-technical users. However, from a technical perspective, the software needs to provide comprehensive capabilities as shown in Table 2.6.

2.6.3 Digital Forensic Investigation Perspective / Technical Perspective

In an eDiscovery Insight blog, “What do Wikileaks and E-Discovery have in Common?”, Leehealey (2011) mentions many technical requirements for e-Discovery such as “email threading, data clustering and document relationships, dynamic in a classified spillage / IP audit and PII (credit card) audit”. This is a famous case of investigation. However, there are many technical aspects of digital investigation that make the e-Discovery process complicated. For example, an analysis and findings report illustrates “data acquisition tasks” where legal experts did not agree with technical experts (Carlton & Worthley, 2009). This is one reason why the e-Discovery process is more comprehensive from the forensic investigation point of view than traditional discovery. Forensic examiners always prefer the right tool for an investigation process. However, without testing and assessing the capability of a digital forensic tool, it can be difficult to decide on the right solution. The *Buyer’s Guide for IT Professionals* (2010) states that

“the technology should easily fit within the organisation’s existing infrastructure without customising or additional resource burden on the IT department. In addition to seamless integration, a solution must have complete functionality to support each phase of the e-discovery process: identification,

collection, preservation, processing, analysis, review and production” (Guidance Software Inc., 2010, p.2).

2.7 SUMMARY OF E-DISCOVERY ISSUES / PROBLEMS

The stealing and escape of sensitive information through a computer network are major concerns and business risks for any organisation (Qureshi, 2009). The challenge is how to find and analyse data from a live network for forensic purposes and investigation. There are many challenges such as high risk litigation, threat of court sanctions, tight deadlines, lack of an e-Discovery process and legal expertise. Although NIST has established a standardised approach for “General Test Methodology for Computer Forensic Tools”, the complexity of the e-Discovery process and tools is such that standardised frameworks for digital investigation still needs to be built. Many e-Discovery tools are yet to be verified and validated before they can be used in practice. Therefore, standardised e-Discovery digital forensic tool verification and validation procedures are yet to be established (Beckett & Slay, 2007). Another major issue with current digital forensics tools is that the techniques used to analyse the e-Discovery process have cost and time constraints, due to the large volume of data involved in the process.

According to Gould’s statistical research, 97% of all new business information is in electronic and/or digital form, 60-70% of corporate data resides in or is attached to e-mails, 183 billion messages are sent per day and more than 2 million emails are sent every second (Gould, 2008). Matthews (2010), describes three e-Discovery case examples show that in one case, e-Discovery took place simply because of a “likelihood of litigation”. Again, all other cases indicate that the e-Discovery process assists with compliance, record management and litigation.

There are many challenging questions about digital investigation in the e-Discovery process that are still not clear. For example, when, why and how, does e-Discovery take place in an organisation? How much does an e-Discovery process cost? Some of the answers to general frequently asked questions (FAQ) are provided by Wiles (2007). During investigation, a digital forensic investigator is faced with many difficult questions, such as “Is it feasible to make a physical image of hundreds or thousands of hard drives in an electronic discovery effort?”

(Wiles, 2007, p.126). Therefore, “successfully addressing the challenges of e-Discovery requires the integration of a range of contributions – from humans, from technology, and from methodology” (“Project Management and Consulting,” 2010).

The growth of digital technologies being used in businesses increases the complexity and the difficulty of the digital investigation due to the large amount of data involved. For example,

“e-discovery today represents 35% of the total cost of litigation, and companies that fail to produce emails and other electronic content in a timely or appropriate manner face the risk of paying millions of dollars in sanctions and fines, not to mention loss of corporate reputation, lost revenue and embarrassment” (Osterman Research, 2010, p.1).

Lack of a well-organised records management process and effective retention practices put any organisation at risk. However, a pro-active approach that implements key policies, appropriate tools and procedures can significantly reduce the fear of litigation.

The use of e-Discovery for outsourced data is another major issue due to the complications caused by the workings of the global network and the enormous amount of data travelling on the network. Most enterprises need to cope with malicious activities by intruders, viruses and security issues on the network every day (Cisco Systems, 2001). To protect private information travelling on the network (e.g. bank account numbers, usernames and passwords, credit card numbers, etc.) businesses need secure networks. However, there are many incidents where intruders have successfully attacked live networks without storing any information on a hard drive. Examples for such attacks, are DoS (Denial of Service), hacking the password file on the web server, theft of services, internal fraud and disclosure of information. However, “attackers fingerprints remain throughout the network, in firewall logs, IDS/IPS, web proxies, traffic captures and more” (SANS, 2010).

A law review article shows the number of different e-Discovery cases increased compare to the last decade (Willoughby, Hunter, & Antine, 2010).

“Sanctions relating to e-discovery violations have reached courts everywhere and have appeared in all types of cases. The most

common case types are employment 17%, contract 16%, and intellectual property 15.5%. Sanctions for e-discovery violations were also discussed in tort cases 11% and a variety of other types of cases, including civil rights 8.5% and bankruptcy 3%” (Willoughby et al., 2010).

The overall statistics from the Duke Law Journal shows sanctions for e-Discovery violations by the numbers (Willoughby et al., 2010).

Richard, Roussev and Marziale (2007) describe a few issues related to using current methodology and software tools, such as digital examiners being confused while extracting evidence. For example, different e-Discovery software tools provide auditing information in a variety of formats. There are limited capabilities of e-Discovery software tools, most forensic tool suites are only designed as general-purpose tools (e.g. keyword searches, indexing and file type identification). Another issue is that digital forensic tools consistently contain implementation errors (bugs) and/or are based on flawed assumptions. Therefore, “...both investigators and the tools they use are prone to errors and this can lead to challenges of the results” (Richard et al., 2007, p.89).

A digital forensic process is the process of e-Discovery for the investigation of files, meta-data and suspicious activities, and includes different sources of electronic communication such as email, online-chat and the transferal electronic documents as email attachments using the internet. “Any and/or all of that data could become subject to e-discovery” (Mathias, 2007, p.44). In addition, improper IT resource use creates many retention issues. “Searching through it all would be a great burden on IT” (Mathias, 2007, p.44). However, Barker et al. (2008) explore the new rules for electronic discovery and how those rules should be reflected in management policies, strategies and IT department for ESI.

Several issues regarding e-Discovery digital forensics have been identified and discussed in Chapter 2. A summary of the key issues and problems is presented in this section to provide a snapshot of the current trends in e-Discovery forensics. These problems are to be further explored in Chapter 3 in order to select a researchable problem that is both feasible and relevant.

2.8 CONCLUSION

This chapter presents a review of the main concepts related to e-Discovery. An overview of commercial e-Discovery software tools has been developed. It describes the major key features, and benefits of some industry-leading software tools (*EnCase*, *FTK* and *Intella*). The overview of EDRM framework identifies the different stages of the e-Discovery digital investigation process and how they are presented in the EDRM. It also presents a brief explanation of each EDRM stage which helps to understand the importance of the framework for the e-Discovery forensic process and why it has been accepted by many large organisations.

The review also covers e-Discovery software tools capabilities and expected outputs. In addition, the general summary of e-Discovery issues has been discussed in this chapter. In order to study and test the reliability of the selected e-Discovery tools, six relevant articles are to be reviewed and studied to find out how other researchers conducted similar research.

In the next chapter (Chapter 3), the research methodology will be discussed in detail along with the research design, data requirements, data processing, data analysis and limitations of the research.

Chapter 3

Research Methodology

3.0 INTRODUCTION

Chapter 2 reviewed literature relevant to the concept of e-Discovery and the significance of the EDRM to the e-Discovery investigation process. The literature has also identified key features of selected e-Discovery software tools. Chapter 3 aims to define a research methodology for this study. An appropriate methodology for assessing e-Discovery tools has to be defined so that it can be used to research the problems and issues raised in Section 2.7. The proposal is to review similar studies and establish the approaches other researchers have used for researching tools for e-Discovery. The findings of similar studies can also provide vital information about their adopted methodology, the tools and techniques they applied and the recommendations that resulted from their research. It is expected that the cost and availability of e-Discovery tools will restrict the laboratory investigation to one of the leading brands.

Section 3.1 reviews six similar studies that relate to e-Discovery investigation approaches, tools and techniques. This review will help identify research processes that work effectively and can be adopted as part of the research methodology in this study. Section 3.2 discusses the research design and includes a review of the problem area defined in Chapter 2, defines the research question and draws up a research plan and a data map. The scope of this research is to assess the capability of a few e-Discovery tools through testing scenarios. However, as part of the research, the capabilities of the tools must first be evaluated. Section 3.2.5 identifies four research phases: assessing the scope of the selected tool /software, testing the performance of the e-Discovery tool, assessing the capability of the selected e-Discovery tool and providing recommendations for further research by forensic professionals. Data will be collected in the form of email, email attachments files, data files including those produced in Microsoft Office, Word, Excel and Power Point.

3.1 REVIEW OF SIMILAR STUDIES

A number of similar studies related to the process of e-Discovery investigation have been reviewed as part of this study. However, in order to select the appropriate methodology for this research, at least five or six relevant studies need to be critically reviewed in order to learn how other researchers have developed their strategies and methodologies in the areas related to the proposed research. Chapter 2 focuses on the definition of, some basic concepts related to e-Discovery and a model and strategy that might provide the best results. This can help identify potential e-Discovery issues that need further research and appropriate techniques to carry out the research.

3.1.1 Digital Forensics: Validation and Verification in a Dynamic Work Environment

The first reviewed publication is by Beckett and Slay (2007, p.1) which states:

“The issues of validation and verification in the accreditation environment and propose a paradigm that will reduce the time and expense required to validate and verify forensic software tools”.

The paper identifies some of the current validation and verification issues that arise due to the diversity of tools and the inability of any individual tool to meet all requirements for an investigation. In addition, this paper discusses a current scientific environment approach using the scientific method as a verification testing tool. The National Institute of Justice survey illustrated that half of the project did not succeed due to a lack of forensically-sound training and a digital evidence policy. Law enforcement requires the meeting of minimum standards for this scientific discipline.

Beckett and Slay (2007, p.2) argue that although there are a number of research hypotheses related to forensic concepts and many frameworks provide theoretical guidelines, none of these specifically discuss the validation of tools and processes. Beckett and Slay (2007, p.5) introduce a new “model of tool neutral testing” for the validation and verification of forensic testing tools that can be used to produce valid results. In most cases forensic tools are quite complex and provide many specific functions from which only a few are used by an

examiner. However, the proposed “model of tool neutral testing” includes the features of extensibility, tool neutrality, tool version neutrality and transparency (Beckett & Slay, 2007, p.5). Therefore, when any tool is tested, a set of metrics can also be derived to determine the fundamental scientific measurements of accuracy and precision. The proposed study uses two examples to illustrate the proposed methods. From these examples, Beckett and Slay (2007) verify the results to determine the validity of the forensic tool. The most common function that every forensic examiner uses in digital forensics is keyword searching. Table 3.1 presents the searching criteria, description and keyword examples.

Table 3.1: Types of Searching Criteria for Keyword-searching (Compiled from Beckett & Slay, 2007, p.6)

Search Criteria	Description	Example
Case Sensitive	Is it upper or lower case or does it matter?	KeYworD
Fragmentation	Fragmented location in disk space.	raw disk level
Compound sentence	Is the keyword surrounded by characters or white space?	-
Compound container	Is the keyword located in a container?	zip or compressed files
Deleted	Is the keyword in a file that is deleted or not?	-
Unallocated space	Check whether the keyword is located in an allocated region of the disk.	-
Slack space	Is the keyword residing at the end of a file or sector?	-
Alternate data stream	Is the keyword in an alternate data stream?	-
Metadata	Is the keyword located in the metadata of a file or disk?	Dictionary entry such as FAT, NTFS, MFT entry or other extension.

Table 3.1 indicates the basic requirements that must be considered when testing a tool for ‘keyword-searching’.

In another example, two separate methods for keyword searching, “a ‘grep’ search over a raw ‘dd’ image (copy) of a file-system and the use of a forensic tool such as *EnCase* or *FTK* over the same ‘dd’ image of the file-system” are presented (Beckett & Slay, 2007, p.6). In theory both methods should produce the same results for a simple, text-based keyword search. However, scenarios tested by Beckett and Slay (2007) resulted in a range of validation results, when compared with an automated forensic tool.

The last two sections of Beckett and Slay’s paper (2007) describe the “validation and verification top level model” for data preservation and data analysis. This model is useful for accreditation, validation and verification, testing, training and development procedures. Figure 3.1 illustrates two testable classes, that of ‘data preservation’, in terms of validation and verification, has four (4) main subcategories, and data analysis having eight (8) categories, which represent a distinct functional dissection of the discipline agreed by the Scientific Working Group (SWG).

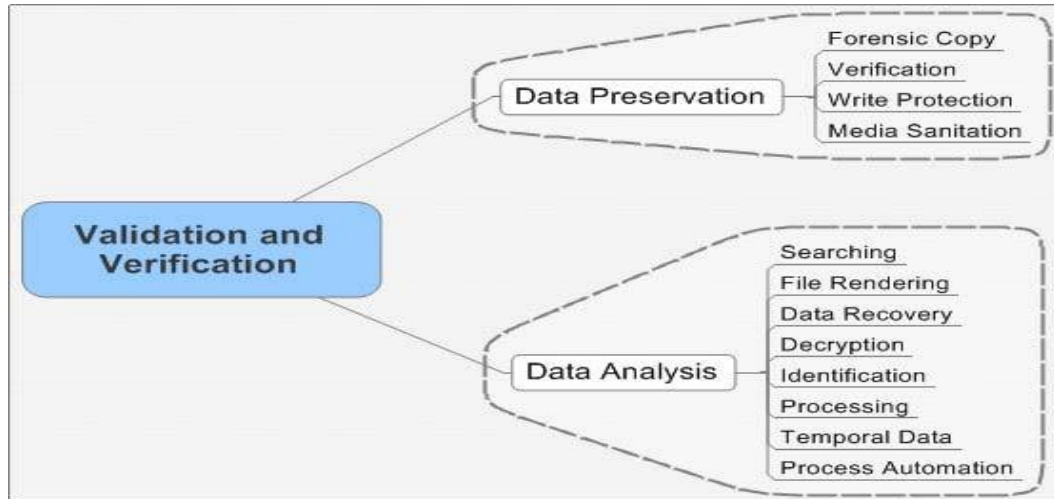


Figure 3.1: Validation and verification top level model (Source: Beckett & Slay, 2007, p.7)

Figure 3.1 explains the validation and verification for ‘data preservation’ testing method for all four (4) sub categories may possibly require different methods and techniques. Similarly, ‘data analysis’ for the ‘searching’ function can be further classified into sub categories for a variety of testable functions, such as Boolean searches, indexing and other searches as illustrated in Figure 3.2.

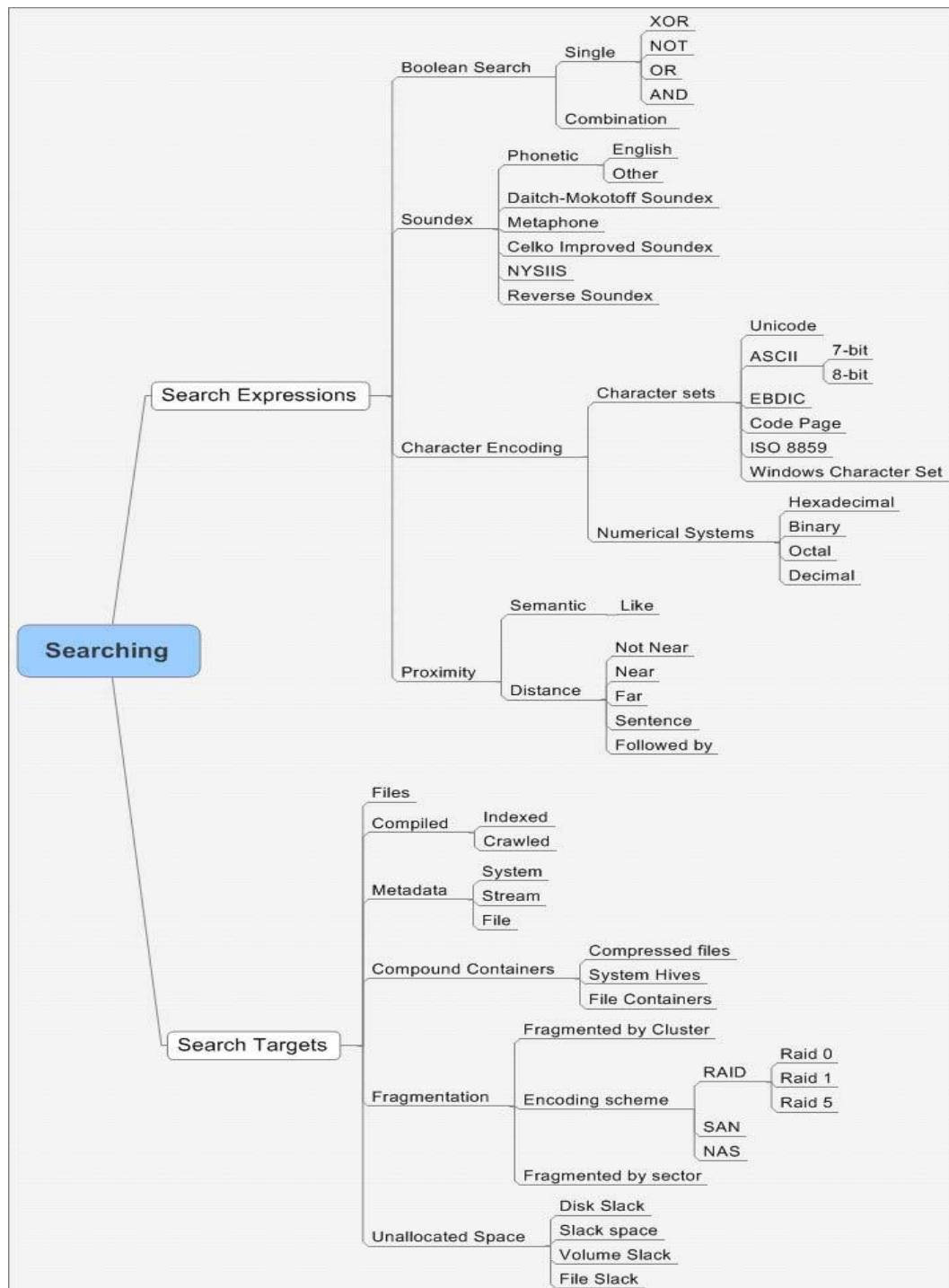


Figure 3.2: Searching breakdown (Source: Beckett & Slay, 2007, p.8)

In the final section of Beckett and Slay's paper (2007), the authors explained the benefit of a reference set that accurately reflects the specifications of the functions in a variety of measures that allows for better precision in determining the validity of a tool. In brief, the proposed paper identifies a credible paradigm in validation and verification in order to assist in the process of digital forensics in a dynamic environment in order to meet the key requirements of accreditation.

3.1.2 e-Discovery Support Tool Design and Implementation of the AGENT Module

According to Kim, Lee, Goo, and Shin (2011), since December 2006, the Federal Rules of Civil Procedure (FRCP) amendments of the United States makes e-Discovery a compulsory requirement for collecting evidence and providing information for litigation. Therefore, every organisation should establish a policy for litigation purposes. For this reason, the proposed e-Discovery support tool (AGENT module) is designed for analysing ESI for e-Discovery requirements (Kim et al., 2011). The overall design of the AGENT module consists of three elements; AGENT, SERVER and MANAGER and are based on EDRM. The purpose of this study is to build an e-Discovery support solution by designing an AGENT module using an EDRM framework. The EDRM guide was used to understand the basic functions, such as search with index, classification, review and analysis of an e-Discovery support tool. Figure 3.3 illustrates the AGENT system configuration, giving an overall idea of how these three modules perform individual roles, and overall communication via a secure network.

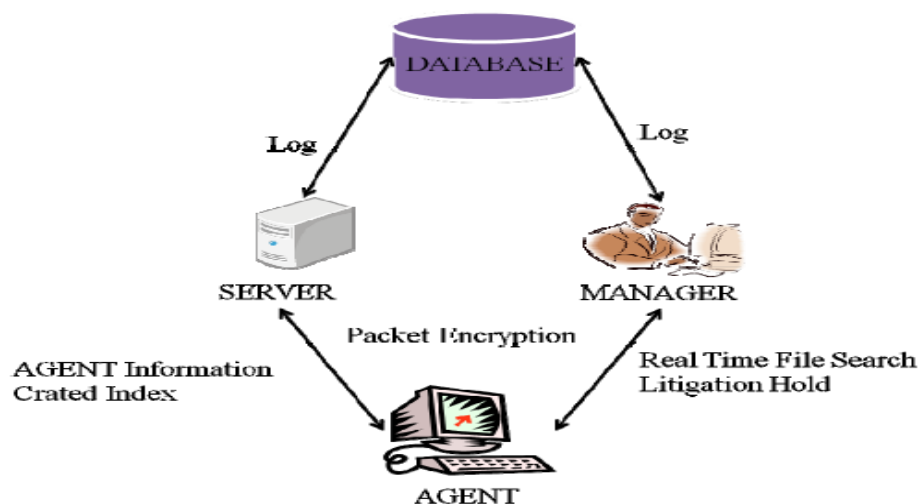


Figure 3.3 AGENT total function configuration (Source: Kim et al., 2011, p.540)

During litigation, the AGENT collects ESI and creates ESI indexes. These indexes contain a lot of significant information such as signatures, keywords, extension and summary of reports that include metadata and indexed results with a time stamp, and are sent to the MANAGER. Once, all information is collected and classified, it is stored in ESI Storage. The MANAGER's role is to review and

analyse the collected ESI. The AGENT acts as a protector and preserve of the ESI, as well as performing the direct, real time file search for e-Discovery if it is required during the litigation hold. The overall system works in a loop, performed by the AGENT in order to process the ESI.

The proposed e-Discovery design and implementation of the AGENT module provides the fundamental e-Discovery process and supported tools using the EDRM set of frameworks. However, there are many research questions raised by this study such as; how successfully does this model work for real e-Discovery? Has there been a real case conducted with this design tool? How does the AGENT collect and index all information? Are there any third party forensic software tools and techniques used for forensic analysis?

The above questions should be taken into account when investigating the process through e-Discovery. In addition, the implementation and the testing of the design tool is yet to be finalised. Therefore, Kim et al., (2011, p.541) state "...future study will expand the proposed model by analysing the requirement of e-Discovery support tool with the implementation of overall design and modules". Theoretically, the proposed design provides much information about the basic e-Discovery processes and its design implementation through database and server. However there are many practical tests required to assess the capabilities of the proposed design.

3.1.3 FORZA – Digital Forensics Investigation Framework That Incorporate Legal Issues

The third reviewed publication by Jeong (2006) discussed FORZA (FORensics ZAchman framework), that incorporates legal issues. Jeong's (2006) paper emphasized the fundamental principal of digital forensic investigation that provides the answers to six basic questions; what, why, how, who, where and when? Jeong (2006, p.S30) stated and recognised the standard procedure of digital forensic investigation and argued that some of the standard procedure is still not aligned properly.

"As many of these procedures were developed for tackling different technology used in the inspected device, when underlying technology of the target device changes, new procedures has to be developed" (Jeong, 2006, p.S30).

Ieong (2006, p.S31) discusses the core principle of “IT security fundamentals” (Integrity, Confidentiality, and Availability) are essential and are mandatory requirements in relation to digital forensic investigation.

Ieong (2006, p.S31) states that “digital forensics investigation is not a static process. Depending on the business nature, system design and legal advice, different methods of investigation would be formulated”.

Ieong (2006, p.S32) discusses the FORZA framework with legal aspects in relation to digital forensic investigation.

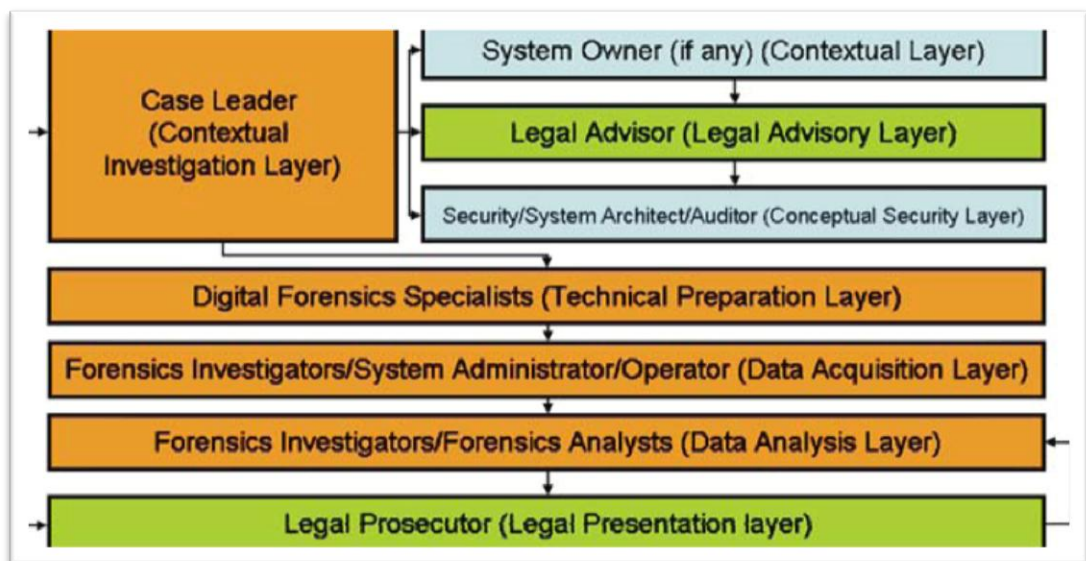


Figure 3.4: Process flow between the roles in digital forensics investigation (Ieong, 2006, p.S32)

Figure 3.4 illustrates the digital forensic investigation process in step by step layers. This indicates the roles and responsibilities of a case leader, as well as the communication between all the relevant participants (layers), including the digital forensics examiner and the legal prosecutor.

Ieong (2006) discusses the FORZA framework, in order to understand how each layer works and its importance to digital investigation in a hypothetical web hacking case. In addition, it recognised that digital forensic procedures are linked together. It shows how different layers are interconnected to each other that integrate into the digital forensic process. Table 3.2 briefly presents the procedures/functions (the how) supporting each layer.

Table 3.2: A high-level view of the FORZA framework (Compiled from Jeong, 2006, p.S33).

Layer	How (the procedures / function)
Case Leader (Contextual investigation Layer)	Request initial investigation
System owner (contextual layer)	Business and system process model
Legal advisor (legal advisory layer)	Legal procedures for further investigation
Security / auditor (conceptual security layer)	Security mechanisms
Digital forensics specialists	Forensic strategy design
Forensic investigators (data acquisition layer)	Forensic acquisition / seizure procedures
Forensic investigators (data analysis layer)	Forensic analysis procedures
Legal Prosecutor (legal presentation layer)	Legal presentation procedures

In brief, the FORZA framework focuses more on digital forensic investigation rather than e-Discovery. However, many layers and roles remain the same for the investigation process. The purpose of reviewing this publication is to understand investigation procedures carried out successfully using the FORZA framework and to understand the methodology of handling a case. In order to differentiate between the digital forensic and e-Discovery investigation processes, Chisholm (2010, p.12) explains the core difference between Forensic Investigation versus e-Discovery in *Integrating Forensic Investigation Methodology into e-Discovery*. According to Chisholm (2010), the scope of the work is the primary difference between forensic investigation and the e-Discovery process. The FORZA framework is more related to typical forensic investigation methodology that involves data acquisition, analysis and reporting.

“In contrast, the forensic analyst’s involvement in the e-Discovery process will likely be limited to technical consultation and the preservation / collection of relevant information. This role may begin as early as the 26(f) conference, assisting legal counsel with technical issues and determining the scope of relevant systems and data” (Chisholm 2010, p.13).

3.1.4 E-Discovery: Identifying and Mitigating Security Risks during Litigation

The fourth reviewed publication by Heikkila (2008) discussed factors important to identifying and mitigating security risks during litigation. An increasing security risk, it is important to protect all ESI with a consistent approach. This paper explains how a company can preserve and protect ESI safely while conducting its daily business during the litigation hold. ESI must not be updated or deleted during litigation which may have a potential impact on the business or may lead to losing the case. As above in Section 3.1.3, Heikkila (2008, p.21) mentioned the core IT security domains (confidentiality, integrity and availability) that control the security risks in producing ESI.

“When producing electronically stored information (ESI) in response to lawsuits, business faces several security risks as well as legal requirements they must satisfy. Customised document management programs and e-Discovery policies are key tools in protecting against inadvertent disclosure as well as meeting business and legal needs” (Heikkila, 2008, p.20).

During the initial phase of e-Discovery, attorneys need to know the locations of their clients’ responsive ESI for litigation purposes, therefore it is necessary to identify the correct location, size, time and other details of the ESI. Heikkila (2008, p.22) illustrates with a visual “Network Map” in (his/her) Figure 1, that indicates the possible location of ESI which assists legal counsel to explain and illustrate overall findings. However, the possible location of ESI varies case to case depending on the legal requirements. For an organisation, a network map is a vital document that gives key information about the entire IT structure of the company. It provides information such as; routers, firewalls, servers (database, exchange, file, email, intranet, and web), information about their IP address, and connected workstations, to identify where possible responsive ESI resides. According to Heikkila (2008), by providing such a snapshot of their network, the company can help the court understand ESI production’s magnitude and complexity. However, the paper did not mention how and from where to collect information from the network, and/or which tool required for collecting ESI. Therefore, it indicates that the illustrated network map was used as a more general

way of understanding where ESI may possibly reside. Heikkila's (2008, p.23) Figure 2, illustrates the legal hold repository for preserving documents identifying network connections to protect documents in order to mitigate security risk during litigation. It includes databases, archives, email, instant messages, backup tapes and all important electronic devices that produce ESI connected to lawsuits.

“A forensics expert should store the ESI in the vault, under the company attorney's direction, by creating a mirror image of the data to save its metadata” (Heikkila, 2008, p.23).

It explains the identifying mission for critical assets that require protection. A proactive approach towards an e-Discovery plan makes it easier to locate all relevant information for a case, project, product or custodian (Heikkila, 2008, p. 23).

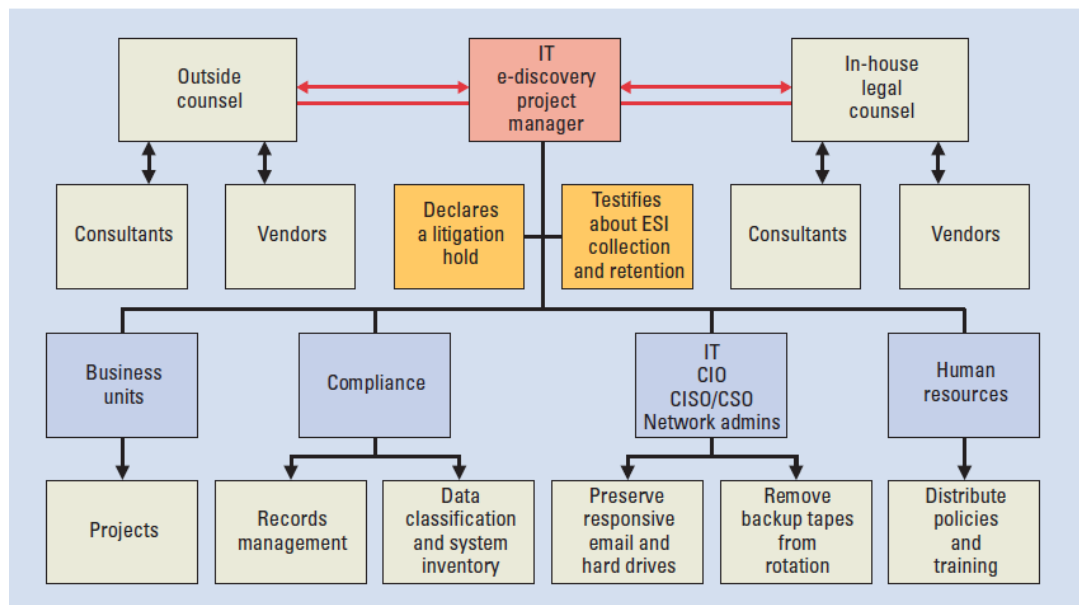


Figure 3.5 E-discovery response team's responsibilities (Source: Heikkila, 2008, p. 24).

The above figure indicates how an IT e-Discovery project manager leads a litigation hold, connected to various departments under legal counsel's direction. The hierarchy of the e-Discovery plan gives an overview of individual department team's responsibilities to identify and mitigate security risk, to preserve sensitive information and critical assets of the company during the litigation process. According to Heikkila (2008, p.24), the e-Discovery project manager needs to fully understand the e-Discovery requirements and IT regulatory-compliance issues. In addition, the e-Discovery project manager also needs to recognise the

legal hold process that can be enforced and how document management programs work to retrieve ESI. An e-Discovery response team is then developed and acts as an incident response team declaring the legal hold. The designated staff control the ESI and investigate where it resides in response to document requests. In some processes, it requires the training of employees in responsive ESI when a legal hold is declared. The individual employee must understand the necessary procedure for an effective implementation of a legal hold. Appropriate data-classification and employee participation play an important role in identifying and mitigating risk for legal requirements and maintaining business policies. The last part of Heikkila's 2008 publication mentions the forensic investigator's roles and responsibilities for collecting and preserving evidence which may be sensitive information. Heikkila (2008) explains how to establish a good document-management program that makes it an easy process to identify and handle security breach issues. In addition, the e-Discovery plan as a proactive approach towards legal requirements controls the risk of security breaches. In summary, Heikkila (2008) provides a lot of information regarding identifying ESI in different locations, company structures, the IT-Network map, policies and e-Discovery frameworks. This information provides a good start for an investigator, designated lead person (e-Discovery project manager), legal counsel or the court in mitigating security risks during litigation. However, there are many technical approaches that require in-depth processes to preserve ESI that were not clarified.

3.1.5 Determining Culpability in Investigations of Malicious E-mail Dissemination within the Organisation

The fifth reviewed publication by Haggerty, Taylor and Gresty (2008) presents an investigation into malicious e-mail dissemination within an organisation to determine levels of culpability within a group.

“Culpability in this sense is the identification of roles that actors played individually and within the whole network to facilitate and encourage the spread of malicious e-mail within an organisation” (Haggerty et al., 2008, p.13).

The conducted study by Haggerty et al., (2008) explained how link analysis is concerned with tying various incidents together during the investigation process, for example, finding information about a malicious image attached to an email.

The digital examiner needs to examine many sources of information such as the file creation date, modification and access on the hard drive, email server logs, server and firewall logs and the file system. More importantly is the location of device information which can be difficult to prove if the evidence resides outside of the jurisdiction of the investigation body. However, in the corporate environment, internal search, seizure and warrants can more easily be based on findings through this link analysis. The link analysis does not necessarily provide evidence admissible in a court of law due to the analytical process, however the techniques of investigation can be useful for a primary source of evidence (Haggerty et al., 2008).

Haggerty et al., (2008) discuss in the context of a corporate e-mail investigation the use of industry-leading forensics toolkits (*FTK & EnCase*) to recreate files and data from a suspect's computer.

“However, these tools do not provide a visualisation of the importance of actors within the social network. In addition, partial data and textual context analysis must be performed manually” (Haggerty et al., 2008, p.13).

This indicates that in many cases even industry-leading software is unable to perform a complete investigation. Therefore, other approaches are required which are forensically sound and valid in court. Haggerty et al., (2008) combines the use of link analysis and network diagrams to identify the route that a malicious e-mail takes through an organisation. The main purpose of using this methodology is to identify source information about the individual responsible for broadcasting the e-mail to other individuals. For an organisational e-mail investigation the suggested methodology is quite useful for identifying breaches of security or usage. Internal policies can focus mainly on employees' communication, behaviour and internal communication via e-mail, attachments and instant messaging. The proposed methodology has participants divided into three categories: *victim*, *passive* and *active*. “Figures 1-4 provide a simple demonstration of application of network diagrams to investigations into the dissemination of malicious e-mails” (Haggerty et al., 2008, p.15).

In order to determine an appropriate level of disciplinary action based on link analysis a level of culpability is determined using a point system (score) illustrated in Table 1 (Haggerty et al., 2008, p.16). In a final step Haggerty et al.,

(2008, p.18) successfully demonstrated victims and active and passive participants using the social network approach.

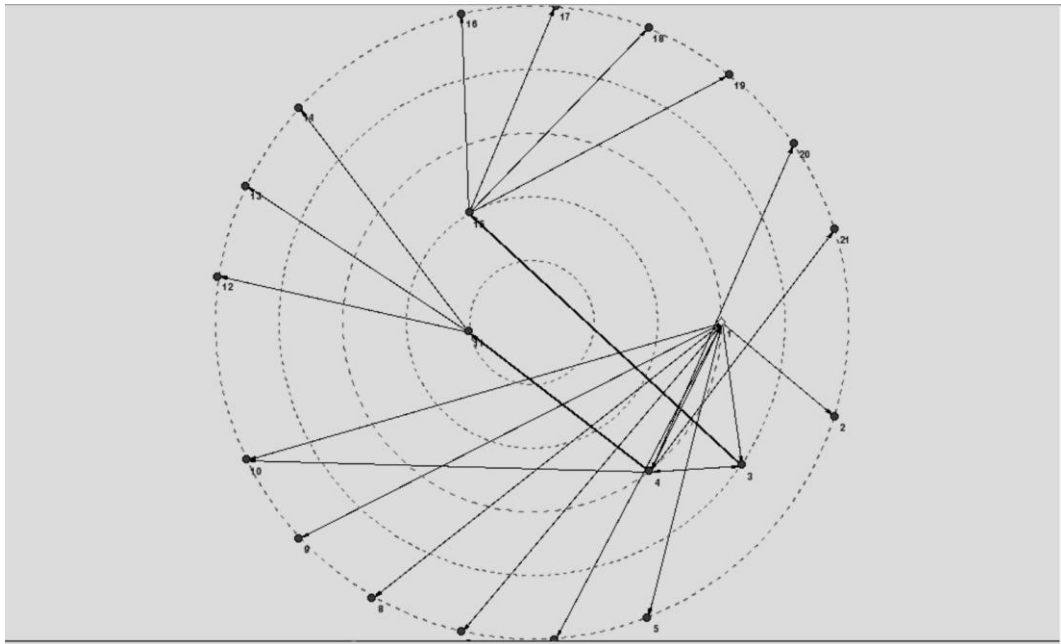


Figure 3.6: Network view demonstrating closeness (Source: Haggerty et al., 2008, p. 19).

Figure 3.6 shows six circles referring to six different levels. The various points (numeric figures) in each circle (level) illustrate scores for determining culpability in malicious e-mail dissemination and the key actors' action as identified by closeness and level of centrality. According to Haggerty et al., (2008) the utility of the tool demonstrates the complexity of identifying the level of centrality; however it provides useful information in identifying the key actors in malicious e-mail propagation.

Overall, Haggerty et al., (2008) discussed the environment of corporate e-mail investigation, a methodology for investigating malicious e-mail dissemination using social network analysis tools in a scientific manner. The proposed approach is good for a primary e-mail investigation process, however, many research questions arise such as how much time does it take, what are the cost and manpower requirements for an overall investigation? How appropriate is it for legal hold and e-Discovery analysis? Has there been any real-world implementation of this scientific approach? The Haggerty et al., (2008) study lead to many further research questions, however this study does teach a primary investigation process for e-Discovery in a digital forensic e-mail investigation.

3.1.6 A Function Oriented Methodology - Searching Function

The last (sixth) reviewed publication by Guo, Slay and Beckett (2009) discussed a function-oriented methodology to validate and verify computer forensic software tools. It presents a scientific and systemic explanation of the Electronic Evidence (EE) discipline through mapping a functionality-oriented paradigm for the digital investigation process. Guo, Slay and Beckett's (2009) research was based on a previous work (Beckett & Slay, 2007), however, in this study the authors concentrated mainly on several distinct functional categories and sub categories of the searching procedure with in-depth analysis. Guo, Slay and Beckett (2009, p.S13) briefly discuss ISO 17025 standards for software validation and verification for trustworthiness of digital evidence, laboratory accreditation, and existing works on EE tool validation and verification. According to NIST, the standardised approach for "General Test Methodology for Computer Forensic Tools" since 2001 and the ISO 17025 Laboratory Accreditation Standard requirements are illustrated in Figure 3.7.

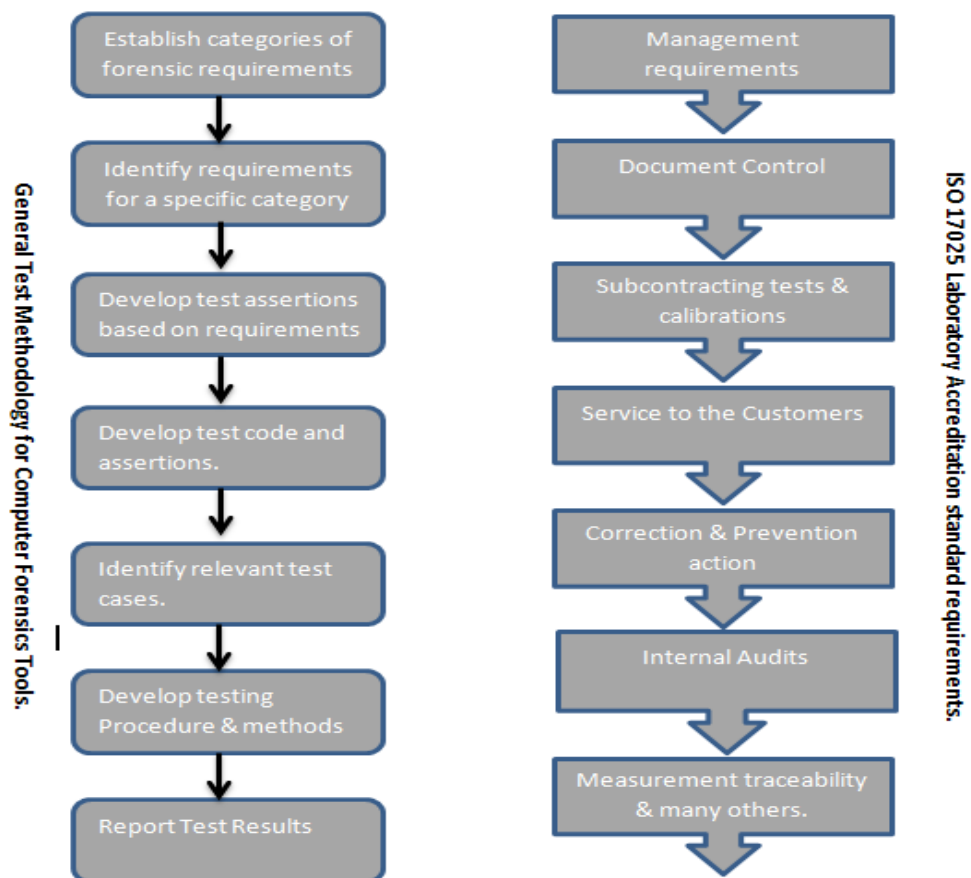


Figure 3.7: Compiled from Test Methodology 7 (NIST, 2001).

In addition, Gou, Slay and Beckett (2009, p.S15) proposed a new functionality-oriented validation and verification paradigm illustrating a scientific and systematic approach, examining EE for digital forensic investigation that may contain similar processes for assessing the capability of e-Discovery tools as a primary focus. The proposed methodology mainly discussed two major processes. The first was how validation and verification are classified into subcategories for data preservation and data analysis. The second was search function mapping. These two methods are very important for digital forensic and electronic discovery (e-Discovery) investigation processes.

Overall, Gou, Slay and Beckett (2009) discussed the necessity for EE tool validation and verification processes when compared to traditional EE tools testing, the validation and verification categories and detailed searching function mapping techniques.

3.2 RESEARCH DESIGN

In the previous Section 3.1, reviews of six similar studies were analysed in order to develop methods that will be used in the empirical research. An evaluation of the related studies will be discussed (Section 3.2.1) and a review of the problem areas from Section 2.7 will be discussed in Section 3.2.2.

The following sections (Section 3.2.1 and Section 3.2.2) will help formulate the research question of Section 3.2.3, which is followed by the hypotheses (Section 3.2.4). The research plan that comprises four main research phases will be formulated based on the main research question and presented in Figure 3.8. Data mapping of the main research question (Section 3.2.6) relating to the research sub-questions, hypotheses (Section 3.2.4), phases of research (Section 3.2.5), and data collection will be presented in Section 3.2.6. The collected data will be analysed and results of the analysis will be linked to the hypotheses that are correlated to the research sub-questions. The answers to these research sub-questions will provide the solution to the main research question.

3.2.1 Evaluation of Similar Studies

Section 3.1 six similar studies were reviewed to identify the standard approach and other, potential methods of assessing e-Discovery digital forensics. Beckett and Slay (2007) discussed the validation and verification of forensic software

tools to improve the speed of the e-Discovery process, using a variety of methods including keyword searching breakdown (Figure 3.2). The second reviewed study provided information about the e-Discovery process and implementation of the AGENT system configuration design through the database and server (Kim et al., 2011). Jeong (2006) proposed a FORZA framework that integrates legal issues in order to understand the principal of digital forensic investigation. Likewise, Heikkila (2008) recognises and mitigates important security risks during litigation for e-Discovery that provides a lot of information identifying ESI, policies and an e-Discovery framework that assists in building a proactive approach to e-Discovery. In addition, Haggerty et al., (2008) discussed investigation techniques through link analysis that help to identify key actors during the investigation process. Finally, Guo, Slay and Beckett (2009) discussed a scientific and systematic approach to finding ESI, test methodology, validation and verification, data preservation and data analysis.

The main objective here is to develop e-Discovery methods and test e-Discovery software tool capability based on a test scenario. The essential element of this research is to identify and to execute a few test scenarios on the selected e-Discovery forensic tools based on the defined test requirements.

3.2.2 Review of Problem Areas (from Section 2.7)

The capabilities and the value of e-Discovery software is based on legal hold. Further processing of e-Discovery might be difficult without assessing the tool's capability in this regard. In addition, Beckett and Slay (2007, p.4) argues that

“The vendor validation has been widely undocumented and not proven publicly, except through rhetoric and hearsay on the bulletin boards of individual tool developers such as Guidance Software and Access Data the main players in this domain”.

Another emerging problem is that many forensic investigators are facing challenges due to the dynamic nature of technology. In many cases, digital forensic investigators are relying on software that was not developed for forensic purposes. However, it still generates the results that are required for the investigation, such as cache memory analysis, a variety of chat log viewers and email applications. In addition, e-Discovery, cost and time constraints, due to the large volume of data, are important factors in an e-Discovery investigation,

raising the research question in regard to assessing the capability e-Discovery tools and further recommendations.

Chapter 2, Section 2.7, discussed many e-Discovery issues in general. However, the problem of awareness arises from the reliability, the performance and report based output. Subsequently, it also emphasised the standard techniques used for e-Discovery software and the technical requirements for the e-Discovery process. In addition, an *e-Discovery Journal* article states that “software acquisitions in the e-Discovery market had underperformed” (Murphy, 2012). Moreover, the response from clients suggested that “the software giants are not yet the best landing spots for e-Discovery software vendors” (Murphy, 2012). However, it is difficult to measure the performance, or compare commercial tools, without receiving limited and/or full licences from software giants. Therefore, the main interest of this research is to test the performance of selected e-Discovery tools by assessing the capabilities of those tools. Successful testing using different file formats, email investigation and e-Discovery techniques using an EDRM framework are the main areas of the research.

3.2.3 The Research Question

The reviewed literature in Chapter 2 highlights many key features for individual e-Discovery software. Therefore, the aim of this research is to assess the capability of e-Discovery tools in terms of performance, complexity (difficulty) in processing files and significant information. The fundamental research question will be based on performance to check the competency of e-Discovery tools.

What performance can be expected of e-Discovery tools when extracting evidence?

In order to answer the main research question, a hypothesis and significant sub-questions need to be formulated. Sub-questions can be derived from the relationship between the testing scenarios and the test results. Therefore the sub-questions formulated are set out in Table 3.3.

Table 3.3: Secondary Research Questions.

<i>Secondary Question 1: How quickly can the tool analyse / produce information for e-Discovery?</i>
<i>Secondary Question 2: What is the complexity of the e-Discovery tool?</i>

Secondary Question 3: Does the e-Discovery tool obtain significant information?

3.2.4 Hypotheses

Digital evidence presented in court must be extracted via scientific methods. Therefore, the working hypotheses will be based on analysis of a demonstrative sample of the evidence by accepting, rejecting and modifying the original hypothesis (Volonino & Redpath, 2009). The key features and functionality of industry-leading e-Discovery tools were reviewed in Chapter 2. Therefore, testing the capability of selected tools to answer the research question is required. The hypothesis of the research will be based on assessing the capability of e-Discovery tools and whether they can be compromised by using a few test case scenarios. Table 3.4 displays the hypotheses for each of the secondary questions presented in Table 3.3.

Table 3.4: Secondary Research Questions Associated Hypotheses.

Hypothesis 1 (H1): The e-Discovery tool can be assessed by evaluating performance and speed of results.
Hypothesis 2 (H2): The capability of e-Discovery software or tools can be measured by testability and observation.
Hypothesis 3 (H3): e-Discovery software features and functionality has speedy / quicker processes for producing results.
Hypothesis 4 (H4): The e-Discovery tool exhibits less difficult / complexity.
Hypothesis 5 (H5): If an open-source e-Discovery tool is used for the same test scenarios, it will be more complex and time-consuming.
Hypothesis 6 (H6): The e-Discovery tool will be able to provide relevant information from the case scenario.

According to the literature reviewed in Section 2.3, e-Discovery software claims specific functionality and key features of individual tools. Therefore, the research

aims to evaluate the selected tools' performance in most successful capability, result from test scenarios.

3.2.5 Research Plan / Phases

The research consists of four phases (Figure 3.8). The first phase is to assess the scope of the selected tool. The scope includes limitations of software tools, availability of e-Discovery software tools and/or licence limitations, e-Discovery software cost and/or budget, document review of the software tools in regard to the objective of the research.

Phase 2 (Figure 3.8) requires Phase 1 to be completed. It involves testing and evaluating the performance of the e-Discovery tool. Phase 3, requires assessing the capability of the selected e-Discovery tool, including a detailed evaluation of the tool based on scenarios and e-Discovery analysis. Finally, Phase 4 provides the presentation and/or a report and recommendations, based on test results collected from the tests and data analysed.

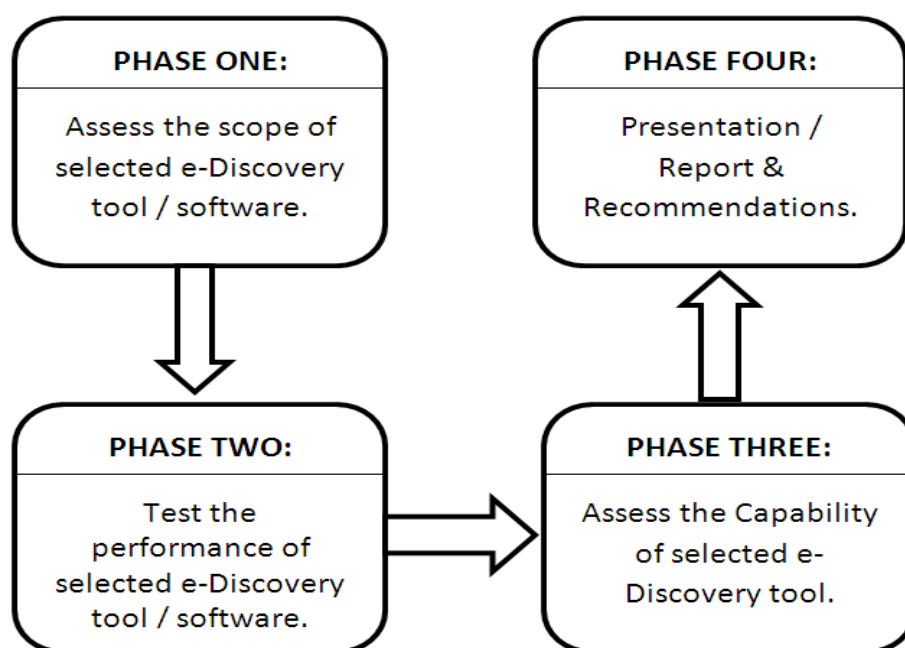


Figure 3.8: Research Phases

3.2.6 Data Map

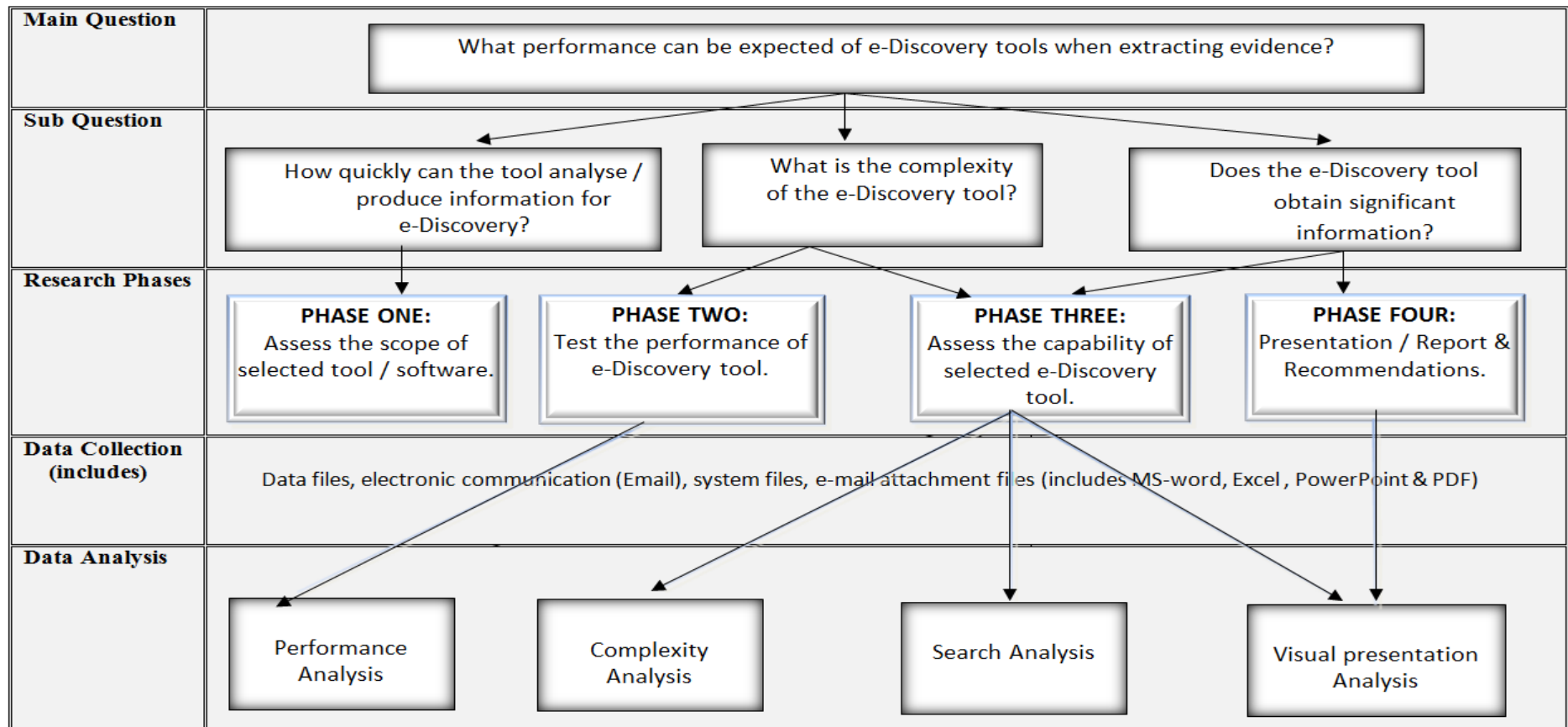


Figure 3.9: Data Map

3.3 DATA REQUIREMENTS

The Enron email dataset is used as the main source of data to assess the capability of e-Discovery software tool (*Intella*) in the field of research.

“Most of the experiments in this field research are performed on synthetic data due to lack of an adequate and real life benchmark” (Shetty & Adibi, 2004, p.1).

However, this dataset is a useful source for research in fields like link analysis, social networking analysis and search analysis. “The Enron email dataset was made public by the Federal Energy Regulatory Commission during its investigation” (Shetty & Adibi, 2004, p.1). The dataset still had a lot of integrity issues and corrupt messages, however it is very useful and similar to a real life data (Shetty & Adibi, 2004).

The information collected in Phase 1 (i.e. assess the scope of selected e-Discovery tools) contains a review of related literature, internet e-Discovery journals, software vendor sources, electronic database resources and document review software tools.

3.3.1 Data Collection

Data collection and preservation is an important part of the digital forensic and e-Discovery investigation process as reviewed in Section 2.4.3 of Chapter 2.

According to FRCP Rule 26 (f) —

“parties must sit down together at least 21 days before holding the scheduling conference to discuss and agree on some form of procedure or protocol to govern the e-discovery process” (Clearwell Systems, n.d.).

Furthermore, the biggest challenges facing collection and preservation implementation is determining what collection methods are required. Each type of data storage requires a different strategy and approach. For example, search and preservation strategies can include inclusive or exclusive searches, file types, dates and times, keywords and metadata. Blumenschein (2011) states that 80% risk is involved during data collection and preservation of ESI. Therefore, data collection requires planning, not only due to the large volume of data, but to minimise risk in a timely and efficient manner. The EDRM Collection Guide,

suggest the following collection methodology for acquiring ESI in litigation matters.

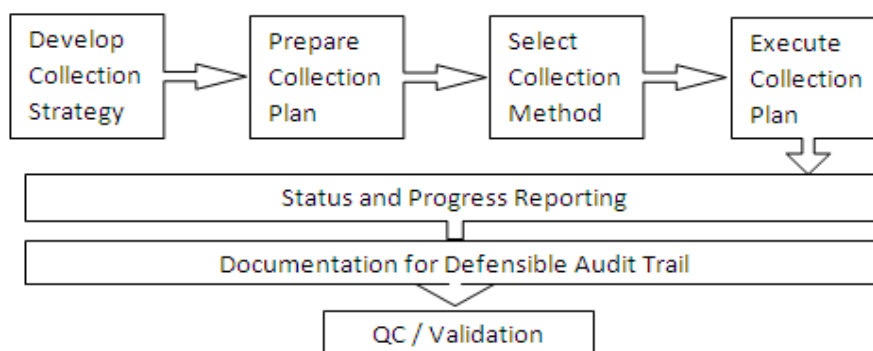


Figure 3.10: Collection Guide (EDRM, n.d.).

“The collection methodology for acquiring ESI in litigation matters, governmental inquiries, and internal investigations in a legally defensible manner” (EDRM, n.d.). Therefore, any organisation entering civil litigation must be prepared to address the digital evidence. Table 3.5 shows major collection sources.

Table 3.5: Collection Source (Compiled from E-Discovery Guideline & Toolkit, EDUCAUSE, 2011)

Types of Data	Collection Source
Data Files	Active, Archive, Backups, Legacy, Internet
System Files	Audit trails, Access controls lists, metadata, logs, Internet “Footprints” such as cookies, internet history and browser activity.
Electronic Communication	Email, Instant Messages (IM), Voice mail
Hardware Devices	Desktops, Servers, Laptops, Personal Digital Assistants (PDA), Mobile Phones, USB Drives, Network appliances, MP3 / IPOD Players, Backup Media (CD,DVD, Tapes) and Internal and external disk drives.
Software Applications	Office suites applications, ERP systems, email systems, CRM Systems, voice mail systems, record management systems, database management systems and other related software.
Other Locations	Work devices, applications and departments, Home devices and applications, third-party devices and applications.

Table 3.5, outlines the wide source for types of data residing in different formats and locations. However, e-Discovery is more concerned with the active and readily available data and not the ambient data that exists in unallocated space

such as, word processing files, spread sheets, e-mail messages, information from database, electronic calendars and contact managers. Therefore, similar to EDRM project data sets, the major source of e-Discovery data collection for this project includes an e-mail data set, a PST data set and a file format data set to fulfil e-Discovery requirements. In general, e-Discovery software provides many methods to analyse results such as de-duplication, keyword searches, hit counts and file extension filters to reduce the number of non-responsive ESI files. However, a few data collection methods adopted in the proposed research are explained in the following sub-sections.

3.3.1.1 Search Function Mapping

Gou, Slay and Beckett (2009, p.S17) mentioned a function-oriented methodology that provides a primary level of keyword searching. Figure 3.11 illustrates “an overview of searching function” that provides different sub categories for specific search analysis such as the searching target, the searching mode and the searching domain. According to Gou, Slay and Beckett (2009) this is the fundamental and scientific approach to a searching query through sub classification.

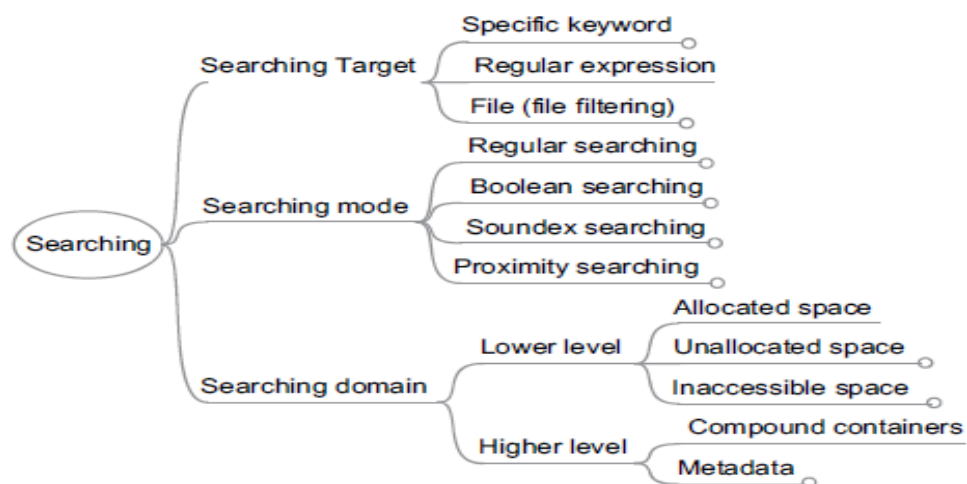


Figure 3.11: An overview of Search Mapping Function (Source: Guo, Slay & Beckett, 2009, p.S17).

Guo et al., (2009) also explained further classification of the search mapping function that illustrates the searching breakdown classification to identify more potential requirements for data collection and preservation.

3.3.1.2 Tiered Custodian Data Collection

Blumenschein (2011) suggests the collection of full disk images, all user-created data, data restriction and the exclusion of known irrelevant file types (applications & dynamic-link library files), keywords, date restriction, known relevant files types (such as Microsoft Office Word, Excel and PowerPoint).

3.3.1.3 Tool Test Requirements

The data sets are the major source of data collection. Therefore, all of the above-mentioned data collection methods should be completed as per the collection strategy. Prior literature will be reviewed to determine the preliminary tool testing requirements. Authors such as Guo, Slay and Beckett (2009) recommend the requirements of NIST's "General Test Methodology for Computer Forensic Tools and ISO 17025 Laboratory Accreditation" should also be reviewed. A number of requirements have been adopted from NIST and EDRM project guidelines reviewed to complete the list of requirements in this research. This research is focused on the active and readily available data mentioned prior with an example (Table 3.5). Therefore, a standalone PC with Windows 7 Professional edition was prepared to install the e-Discovery tool (*Intella 1.5.2*) with minimum configuration requirements.

3.3.1.4 Development of Test Scenarios

The development of the first test scenarios focuses on two validation metrics, namely performance analysis and search analysis. The second test scenarios focuses on another two validation metrics, complexity analysis and visual presentation analysis illustrated in the data- map, research Phase 3 (Figure 3.9).

3.3.1.5 Testing of e-Discovery Tool

The aim is to perform two major tests in Phase 3, according to the test specifications and the assessed capabilities of the selected tool developed in Phase 2. Phase 3 will provide the major part of this research offering important analysis and results. The selected tool will be tested against four major functionality categories; performance, complexity, comprehensiveness and reporting. The selected e-Discovery tool undergoes two test scenarios and each scenario combines test assertions developed in Phase 2.

3.3.2 Data Processing

Data Processing, briefly discussed in Chapter 2, Section 2.4.4, requires identifying, reviewing and processing ESI items as per the project requirement. Here, the aim is to “perform actions on ESI to allow for metadata preservation, itemization, normalisation of format, and data reduction via selection for review” (EDRM, n.d.). For example, data processing requires adopting certain strategies for processing e-Discovery data, such as saving metadata (the complete details of a file), establishing a chain of custody, removal of de-duplication redundancies, search strategy and culling data.

The accurate indexing of a large volume of data is a real challenge. For example, only a few commercial e-Discovery software packages are capable of processing a terabyte of data per day accurately and efficiently. Data processing methods will be implemented to test the results of the e-Discovery tool in the form of indexing, and metadata files generated by the selected e-Discovery tool. The format and the information contained in the data set vary from case to case. Therefore, the results and the related information are collected and summarised into a table, after each test is completed. When all the performance tests are completed, the result of each test will be shown to identify the capability of the tool in each test scenario.

3.3.3 Data Analysis

The “Analysis Phase Diagram”, shown below (Figure 3.12), is the second level framework exposing all the phases of the EDRM Framework under the analysis components explained in detail on the official EDRM website.

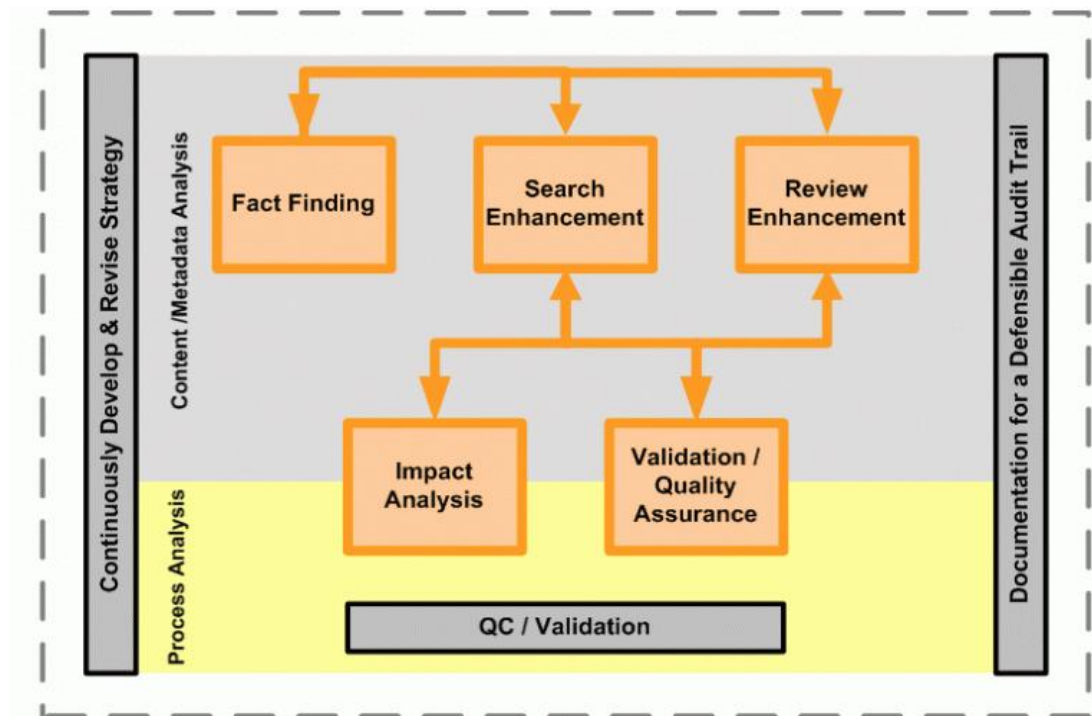


Figure 3.12: Analysis Phase Diagram (Source: EDRM, n.d.).

The analysis phase diagram is used to provide a roadmap for e-Discovery and pre e-Discovery analysis on a broad spectrum. However, data analysis depends upon the specific case and the data set it was gathered from. In Section 3.3.1, appropriate methods applied to data collection that are a major part of the analysis stage were presented. The purpose of data analysis here is to identify software capabilities from data collection. Akers, Mason, and Mansmann, (2011) suggested important key points (Table 3.6) needed to identify the capability of a tool while doing data analysis, however there can be various things to take into consideration while doing data analysis, depending on the requirements of the case.

Table 3.6: Data Analysis Key Points (Compiled from Akers et al., 2011, p.8)

<ul style="list-style-type: none"> • Identify relationships in the content or via metadata analysis.
<ul style="list-style-type: none"> • Identify Consequent metadata includes key phrases, data categories, clusters.
<ul style="list-style-type: none"> • Identify contextual groupings email recipients through metadata analysis.
<ul style="list-style-type: none"> • Identify keyword via search mapping function discussed in section 3.3.1.1.
<ul style="list-style-type: none"> • Identify Indexing

<ul style="list-style-type: none"> • Identify and organise the files related to keyword searches.
<ul style="list-style-type: none"> • Remove duplicate content from responsive document sets.
<ul style="list-style-type: none"> • Identify the versions of data sets.
<ul style="list-style-type: none"> • Identify email conversations through email analysis discussed in section 3.1.5.

By undertaking e-Discovery software analysis, this research will come to understand the strengths and limitations, similarities and differences of selected tools that will give the performance, capability, and complexity analysis illustrated in Section 3.2.6 data map, research Phase 3.

3.3.4 Data Visualisation / Presentation

“If the entire process is considered, ‘From script to screen’...then presentation is the screen in the courtroom. Simply stated, if the material is not displayed properly in front of the jury, then all the effort is for naught” (EDRM, n.d.).

Therefore, presentation is the last and crucial stage of any e-Discovery project. Once major capabilities are identified in data analysis stage, a documentation report can be generated for a defensive audit trial and/or final presentation report for court room use. For instance, this research project produced a report based on data collection, analysis findings, and the testing scenario to satisfy the basic requirements of the research project.

3.4 LIMITATIONS

The proposed research intends to assess the capability and performance of an e-Discovery tool using a few case scenarios from collected data sets. However, certain limitations are expected in the proposed research.

At this stage, it is not believed that any one e-Discovery tool on the market has all of the capabilities and different approaches for e-Discovery discussed above to fulfil the objective of e-Discovery. However, it is best practice to always keep up with the newest of technological advances using standardised approach by NIST as discussed earlier in Section 3.1.6. Commercial tools for e-Discovery analysis is the major focus of investigation, however, it is also difficult to buy

licenced software (tool) for e-Discovery for industry- leading software such as *EnCase e-Discovery*, *AccessData E-Discovery* and *E-Discovery* – Clearwell. There are various factors such as limitation of open-source tools validity, availability and acceptance. In addition, the limited budget and time constraints of this study means the research focuses on understanding and implementing the methodology on one selected tool and test the case scenarios to assess the capability of the tool.

Moreover, a limited set of test case scenarios are designed and tested due to time constraints and the limited licence (trial version but with full functionality). Further testing scenarios could provide dynamic performance by using the various functionalities of the selected tool.

3.5 CONCLUSION

Chapter 3 presented a comprehensive review of similar studies published to understand potential methods and/or different techniques of assessing e-Discovery digital forensic tools, described in Section 3.1. The research design developed in Section 3.2 to review the problem areas of Section 2.7, explained the research question and a data map was drawn according to the requirements illustrated (in Section 3.2.6). Section 3.3 discussed the testing requirements and the data requirement strategy that includes, data collection, data processing, data analysis and visual presentation of data validation and case scenario processing.

After understanding the fundamental requirements of this research study, the next step is to undertake the experimental test scenarios according to the adopted methodology mentioned in this chapter. Therefore the following chapter reports the findings of the empirical research study based on the case scenarios.

Chapter 4

Research Findings

4.0 INTRODUCTION

The literature review in Chapter 2 has defined the term e-Discovery and documented its importance e-Discovery. Chapter 3 defined the research's methodology and laid out the various research phases. It provided guidance in terms of an e-Discovery forensic framework and research approach. The purpose of Chapter 4 is to report the raw findings from the experimental work based on the descriptive methodology outlined in Chapter 3. The raw results from this chapter will report findings based on case scenarios, performance testing, analysis and verification of the e-Discovery tool. The search analysis capability, defensibility and professional understanding of the complexity of the e-Discovery tool are the objectives of this research.

Data collection (data set of e-Discovery testing), processing and analysis proceeded according to the specifications defined in Chapter 3. This chapter reports the discoveries made from the research specifications and how the e-Discovery tool performed in each test case. There also follows a summary of findings collected from the test results and analysis of each test case. Therefore, testing of the e-Discovery tool's capabilities, comprehensiveness and search result precision are an integral part of the analysis.

Chapter 4 is split into four major sections. In section 4.1 the strategy and planning, the case scenarios based on the EDRM framework, the data collection, the processing and the analysis are reported. Section 4.2 provides a report of the findings from the experiment work. Section 4.2 is split into two sub sections, namely, the testing environment and the case scenario findings. Findings are reported in a descriptive manner based on the EDRM framework outlined in Chapter 2. The following section, Section 4.3 is the research analysis based on the findings of Section 4.2. The final section, Section 4.4 provides a conclusion and links to Chapter 5's discussion of findings.

4.1 PRE-PROCESSING PHASE

In order to process the large amount of ESI, the AUT Digital Forensic Laboratory was used to setup the digital forensic workstation, e-Discovery toolkit and other appropriate resources. In the next step, the original dataset was transferred through ‘UltraBlock Forensic USB Write Blocker’ to the AUT Digital Forensic Laboratory for further processing. A forensic copy of the USB drive was taken for pre-processing. The original USB was then placed in an anti-static bag and stored in a secure place for preservation. All further analysis was conducted on the forensic copy. Table 4.1 explains in brief the strategy, planning and the expected outcomes of the pre-processing phase.

Table 4.1: Pre-processing Strategy & Planning

Strategy & Planning	Operational Action	Expected Outcomes
<ul style="list-style-type: none">• Ensure that testing activities will meet the basic requirements and objective of the research.• Identify software integrity, capability to meet the functional requirements, system compliance and interface specifications.• Review and analyse the effectiveness and efficiency of the e-Discovery tool and develop testing strategies as per EDRM framework.• Dataset collection to be based on test case scenarios	<ul style="list-style-type: none">• Evaluate, plan, develop and deploy testing techniques for new case scenarios.• Enhancements to existing case scenarios throughout EDRM framework / strategy.• Create and execute test cases and scenarios to determine optimal system performance according to NIST specifications.• Conduct all types of testing to analyse, process and produce results using accepted testing techniques.	<ul style="list-style-type: none">• Existing system reviewed and equivalent e-Discovery tool tested.• Test the results in light of e-Discovery tool capability, performance and speed.• Test and verify the key features of relevant e-Discovery tools’, functionality.• Analyses of formal test results as per test requirement.• Produce reports for all testing efforts, results, activities, data and findings.

4.1.1 EDRM Framework

Chapter 2, Section 2.4 discussed EDRM Framework guidelines. Both test case scenarios are using the EDRM Framework as a benchmark. Table 4.2 shows the test case scenario segmented into different stages of the EDRM Framework for the case investigation procedure.

Table 4.2: EDRM framework applied to test case scenarios

EDRM Framework / Stages	Description As per EDRM	Potential Source / Procedure / Action
Identification	Locating potential sources of ESI.	Search people, location, files, folder, archives, and external devices.
Preservation	Ensuring that ESI is protected against inappropriate alteration or destruction.	USB stick preserved in anti-static plastic bag and stored in a secure place.
Collection	Gathering ESI for further use in the e-Discovery process.	Data set from EDRM Data set download files, folder and .Pst.
Processing	Reducing the volume of ESI.	Indexing, culling, discussion thread creation and de-duplication
Review	Evaluating ESI for relevance & privilege.	Email, documents, Jpeg, and different file formats.
Analysis	Evaluating ESI for content & context, including key patterns, topics, people & discussion	Keyword searching, link analysis through email investigation, files & folder analysis, de-duplication and comparison of differences.
Production	Delivering ESI to others in appropriate forms & using appropriate delivery mechanisms.	Produced report in .pdf, html and other required formats.
Presentation	Displaying ESI before audiences, especially in native & near-native forms.	Prepared report in appropriate format.

4.1.2 Volume Reduction

In general, volume reduction will increase the processing speed of work for e-Discovery case assessment. KPMG (2010, p.18) illustrates how ESI volume can be reduced by processing and culling.

Table 4.3: Volume Reduction for the e-Discovery Process

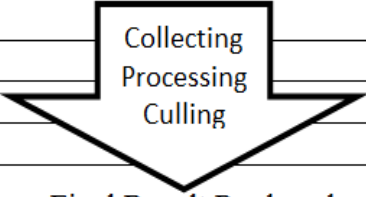
Description	Test Case (TC01)	File size Volume Reduction
(File size – Actual)	1.0 GB	
Data files collected	15,658	
Processed	6,958	
Culling	1,460	
Relevant documents reviewed	204	
		Final Result Produced

Table 4.3 shows an example of the first test case scenario (TC01) that illustrates how actual file size (1 GB) of ESI can be reduced by processing and culling through the relevant information. The numeric figure of the test case (TC01) indicates the file size volume reduction from 15,658 files to 204 files at the end of the process. The methodology used to reduce the volume of test case files is illustrated in Figure 4.1 of Section 4.2.2.1.6.

4.2 FIELD FINDINGS

The field work was carried out on two case studies, illustrated in Table 4.4.

Table 4.4: Case Scenarios

Case Scenario	Description
Case Study TC01	A – Empirical Low Risk Case Scenario – e-Discovery
Case Study TC02	B –Empirical High Risk Case Scenario – e-Discovery

These cases are different; Case Study TC01 was used as a low-risk case scenario for an internal email and ESI investigation of employee e-Discovery. In general, low-risk “...investigations are less likely to result in a formal prosecution, but are more likely to end in disciplinary action” (Haggerty et al., 2008, p.12). Case Study TC02, an empirical high-risk case scenario was used to investigate unusual communications by an employee, distributing secret business information to an external person. High-risk cases are more likely to result in a formal prosecution. The empirical case scenarios were investigated using a powerful e-Discovery search tool named Intella from Vound Software. The aim of this chapter is to assess the capability of the e-Discovery tool through verifying the results and analysing processing speed, performance and complexity. The field findings of the empirical case scenarios are reported in Section 4.2.2.

4.2.1 Testing Environments

Two execution environments were used for testing the e-Discovery software (*Intella 1.5.2*), a standalone computer (Workstation 01) with Microsoft Windows 7 – 32bit professional edition with 4.00 GB installed memory and a wireless laptop computer (Workstation 02) with Microsoft Windows 7 – 32bit ultimate edition with 2.0 GB Installed memory. The hardware specifications are illustrated in Table 4.5.

Table 4.5: Hardware configuration for workstation 01 and workstation 02

Configuration	Workstation 01	Workstation 02 (Laptop)
Processor	AMD 9650 Quad-Core Processor	Intel (R) CPU Dual Core
Processor Speed	2.3GHz	1.73 GHz
Installed Memory (RAM)	4.00 GB	2.0 GB
Operating System	Windows 7 Enterprise	Windows 7 Ultimate
System Type	32-bit O/S	32-bit O/S

For the hardware specification, a “Device Manager” snapshot was used on both operating systems to record accurate system configuration. Test case scenario TC01 and TC02 are the examples that followed the generic procedures.

Table 4.6: Support software (e-Discovery tool) information

Software	Version	Description
Intella 100 GB	1.5.2	“Vound’s innovative email investigation and e-Discovery software is quickly becoming the preferred tool for enterprise, professional service firms, law enforcement, government agencies, and law firms worldwide”(Vound Software Inc., 2011).

Table 4.6 shows brief information about the e-Discovery tool (*Intella 1.5.2*) used to configure and setup both case scenarios.

4.2.2 Test case scenario findings

Section 4.2.2 is divided into two parts one for each case scenario. The first test case scenario’s (TC01) investigation findings are described in sub-section 4.2.2.1 entitled A – Empirical Low-Risk Case Scenario – e-Discovery. The second test

case scenario (TC02) investigation findings are described in sub-section 4.2.2.2 entitled B –Empirical High-Risk Case Scenario – e-Discovery.

4.2.2.1 A – Empirical Low Risk Case Scenario – e-Discovery

Case Study: TC01 Employee distributes inappropriate images via e-mail.

The CEO of XYZ company retained the AUT Forensic Laboratory to investigate e-mail communication by an employee suspected of engaging in the sharing of inappropriate images via the company's internal network that had resulted in a breach of company policy. The company's CEO decided to investigate the case. The case findings, processing, analysis and result employed the EDRM framework as an investigation procedure as discussed earlier in Chapter 2, Section 2.4.

Disclaimer

The chosen case scenarios are for experimental research purposes only. Individual names and digital information presented in the case scenarios are fictitious and are not intended to reflect actual people or places.

4.2.2.1.1 Introduction

The bankruptcy of Enron is the largest, and one of the most striking, corporate collapses in history. Enron was among the most successful companies in the world, established in the Fortune 500's top 10 with estimated sales of US\$ 100 billion per year.

The aim of this scenario is to investigate the test case's digital format email files (.PST format) using the e-Discovery tool. The data-set, in the form of general files and folders, has few PST's archived and downloaded from the EDRM site resources.

4.2.2.1.2 Objective

The XYZ company selected the AUT-Forensic Laboratory for e-Discovery services to process and select complex email messages and attachments from its email files.

Table 4.7: AUT Digital Forensic Laboratory Enron Email Investigation

Action	Description
What	Enron Digital Evidence (.PST file)
Why	Test, search, process, analyse electronic files for e-mail investigation.
Where	AUT Digital Forensic Laboratory
When	2 nd February, 2012.

4.2.2.1.3 Research challenge

Searching 6,958files, 23 folders and 8,587 messages of PST's source of electronic files, decentralised searching (full-text).

Scenario 1 preparing the email investigation for email analyses

4.2.2.1.4 Identification

Table 4.8 shows a brief summary of identified actors, however there are many other employees' emails associated with these findings. For example, Figure 4.3 shows time analysis with linked employees. The complete list and details were provided in an export report. The complete and/or original message hash and name are not illustrated, due to privacy concerns.

Table 4.8: Potential source of ESI (Suspect list and details)

Suspects	Message Type	Message Hash	Sent	Received
Matt	message/rfc822	df510c7db6a0d117507dae...	Jun 13, 2000 22:09:00	Jun 13, 2000 22:09:00
Susie	message/rfc822	598d8343920d090d760fb9...	Jun 03, 2000 08:14:52	Jun 03, 2000 08:21:00

4.2.2.1.5 Preservation and collection

The original data set was transferred to the digital forensic workstations' hard drive storage using a write blocker. The original USB drive was stored in an anti-static evidence bag in a secure place for preservation of evidence. All further analysis was conducted on the forensic copy. According to Ball's (2009) guidelines, all e-mail includes the extensive metadata of the sender and the receiver. The header data contains all the information about email communication from source to destination with timing. The information was preserved in a secure place and a screen shot was taken as a backup. Similarly, other important files and all attachments' original metadata were preserved, a log of the original system's metadata was produced along with the files. In neither of the case scenarios was

there a requirement to build a data map, collecting relevant information from the server, the database and any other places for example, because the test case scenarios were based on the Enron-dataset only. However, building a data map is vital for collections.

4.2.2.1.6 Processing and analysis

In both case scenarios, high priority of investigation was given to electronic communication and/or electronic mail (email) because “Emails comprise both structured and unstructured data. Structured data provides qualitative information to the forensics examiner” (Haggerty, Karran, Lamb & Taylor, 2011, p.15). Unstructured data is more complex, requiring processing and analysis techniques used to identify the key actors, email relationships and other findings. As per the EDRM Framework, the processing stage guidelines are used to reduce the high volume of data using different techniques such as indexing, culling, clustering and de-duplication of the ESI to speed up the process. The first step carried out to reduce the size of email data volume (by finding relevant information as illustrated in Figure 4.1) used a search methodology, processing techniques and the culling of irrelevant document.

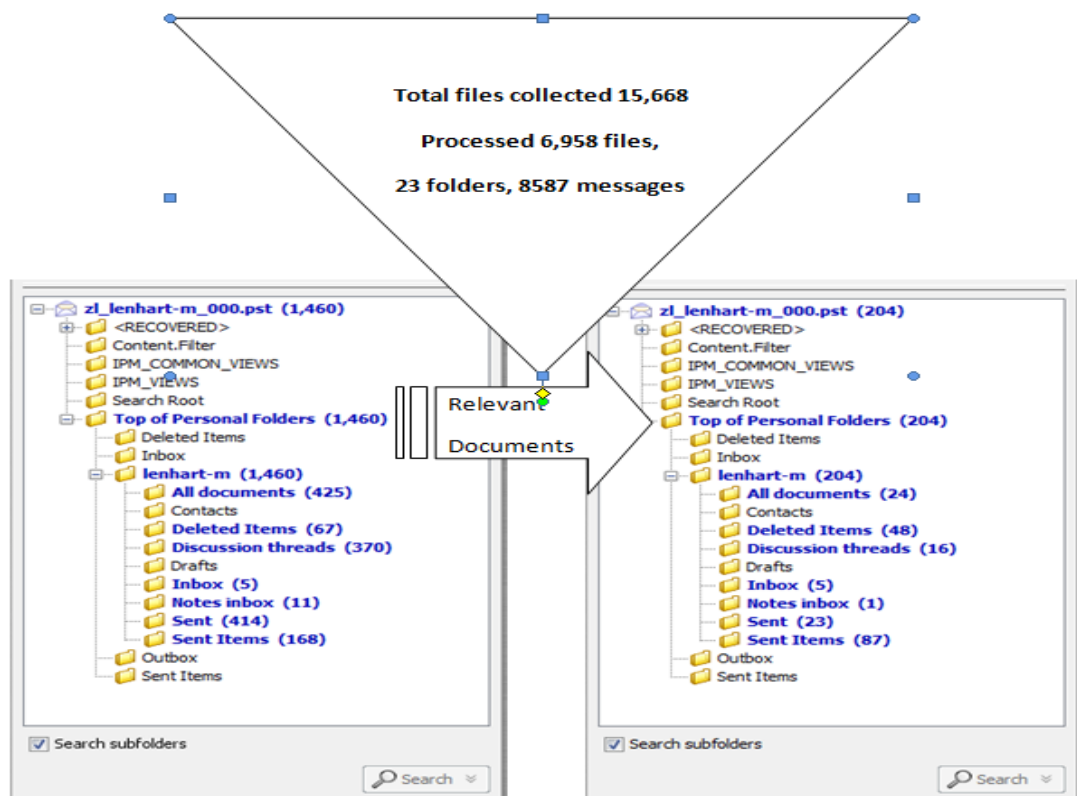


Figure 4.1: Volume reduction of the test case TC01 scenario

The reduced volume of relevant information was further sub-processed using an analytic approach. For example, processing through file system level metadata, application level metadata, structural metadata attributes with “tagging”, cluster associations for similar documents and near-duplicate associations for similar documents.

The next stage is the analysis of the relevant information. With the use of the e-Discovery tool (*Intella 1.5.2*) further analysis was carried out to assess the capabilities of the tool. The analysis of case findings includes: different types of keyword searching such as boolean, fuzzy, proximate and other appropriate types, link analysis through email investigation (identifies the email and instant messaging conversations that occur between parties), files and folder analysis, de-duplicating and comparative differences. The dataset contains the folder information for each of the company’s employees. Each message present in the folders contains the sender’s and receiver’s email address, date and time, subject, body, text and some other information. Figure 4.2 represents the one of the two suspects’ images, music files, video clips and other files attachments shared through email communication.

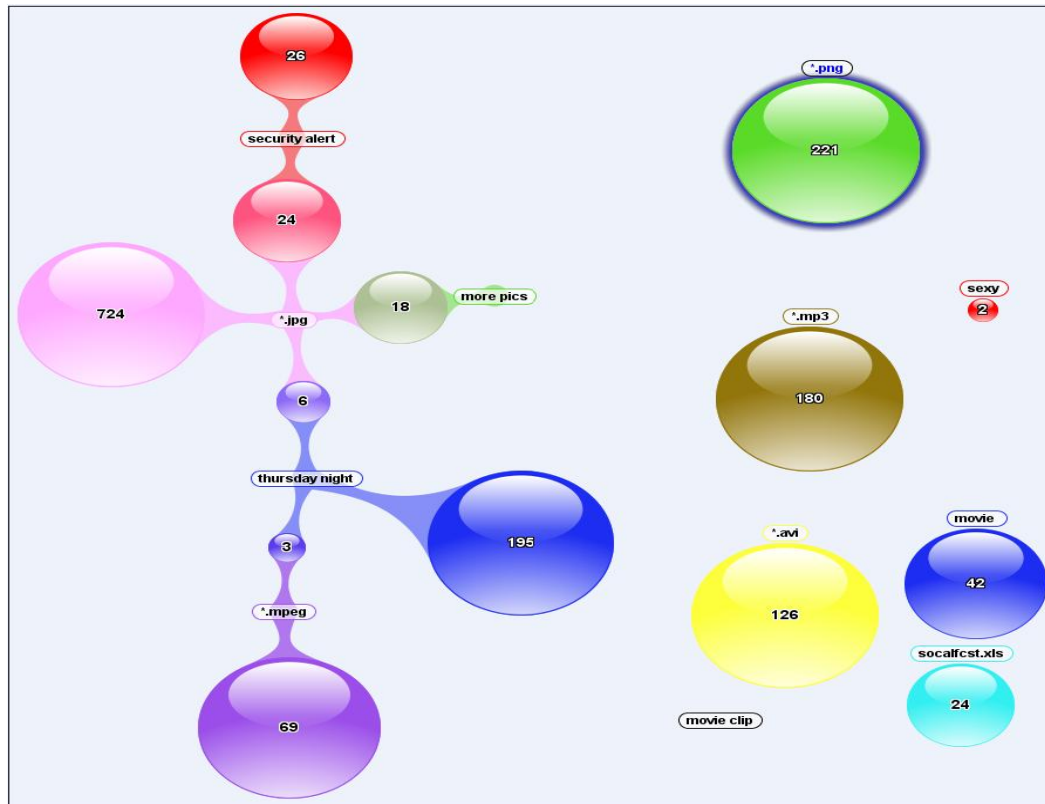


Figure 4.2: The visual presentation of file type search results

The investigation would prove to be more difficult than at first thought, however, the extensive features of the vendor’s e-Discovery tool successfully recovered the deleted messages from the target individual’s personal folder. Another great feature, parent and child relationship of email thread, of the e-Discovery tool *Intella 1.5.2* makes the investigation process able to clearly find its way to the target suspect.

Figure 4.3 illustrates time analysis of the target individual’s email communication. The .PST file contains a lot of private email addresses therefore, original contact details text is manually removed by formatting and hiding the correct details to ensure privacy. However, the time stamp, header information, MD5 value and other important information from the email communication is recorded.

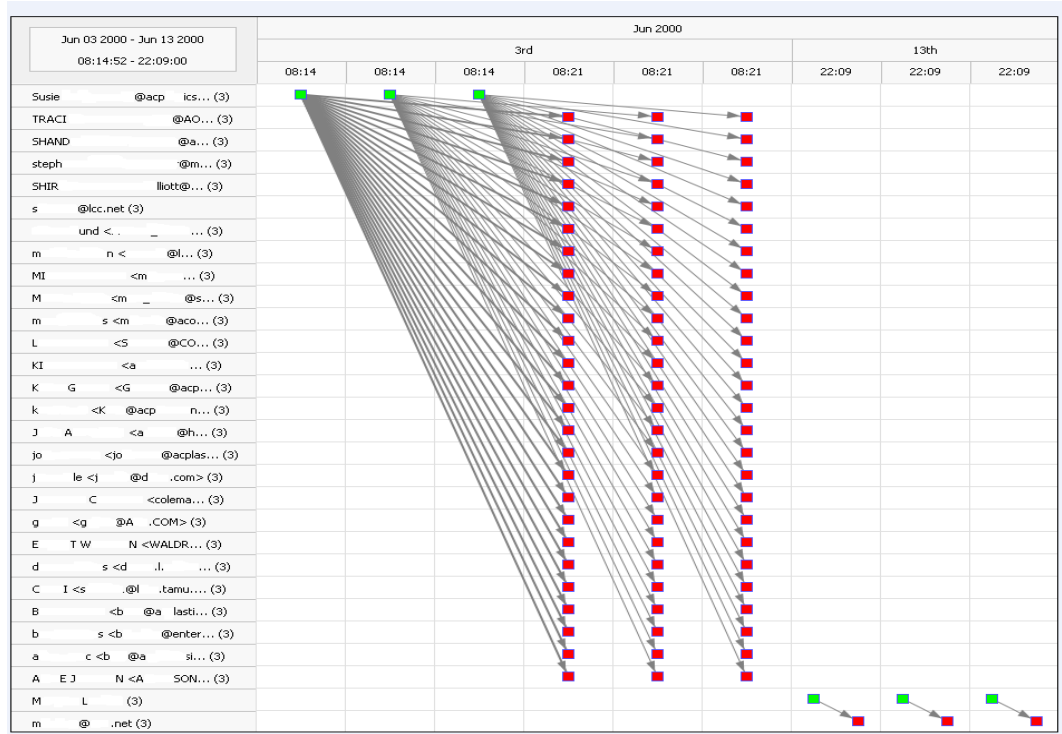


Figure 4.3: Time analysis of target’s email communication

“The Timeline view shows a chronological representation of email communications. The left pane shows the email senders and receivers, with their communication plotted chronologically. Every arrow in the timeline view is an email and points to the receiver of the email. The green squares are senders. The red squares are receivers on the To-list” (Vound Software Inc., 2012, p.147).

Since the files had been compressed they had to be transformed and viewed in a special manner to be able to see the true content.

4.2.2.1.7 Presentation of findings

This section provides the presentation of findings using a template for an examiner's report based on NIJ guidelines for documenting and reporting. Table 4.9 shows the actual findings of Case TC01.

Table 4.9: Case TC01 brief special report compiled from NIJ (2004, p.25)

Report on Email Analysis	
Subject: Email Investigation / e-Discovery for suspect employee & associates.	
Status	Closed
Items Analysed	<ul style="list-style-type: none"> ▪ Tag Number : 01234 ▪ Item Description : Enron Digital Evidence (.PST file)
Tool Used	<i>Intella 1.5.2</i> (Vound Software)
Assessment Processing & Analysis	<ul style="list-style-type: none"> ▪ A forensic copy of the USB drive was taken, the original USB stored in an anti-static evidence bag and stored in a secure place for preservation. ▪ The examined suspect's .PST file was found to contain some explicit images shared through the internal email network. ▪ Deleted files were recovered by <i>Intella 1.5.2</i> ▪ .PST file data, including: attachment file names, dates and times, size and complete path were recorded. ▪ The suspect's email content, graphics files and HTML files were opened, viewed and recorded.
Summary of Findings	<ul style="list-style-type: none"> ▪ 425 document files containing images in different file extension formats (.png, JPEG) were recovered from suspect's personal (email) folder. ▪ 67 deleted files were recovered from suspect's personal (email) folder. ▪ Out of 626 JPEG images, there were 24 JPEG files of explicit images that were linked to the suspect's top personal email folder as inappropriate attachments

	shared with colleges. ■ Summary table of total number of files recovered.			
Statistics for total files recovered	MP3 Audio Clip	138	TIFF Image	3
	WAV Audio Clip	4	WMF Image	50
	HTML Document	39	Internet Location (URL)	4427
	MS Office Document	340	Email Message	8640
	MS Word Document	54	vCard File	12
			MS PowerPoint Document	44
	PDF Document	3	MS Excel Document	326
	Plain Text Document	23		
	MS-DOS/Windows Executable	48	AVI Video Clip	66
	BMP Image	1	MPEG Video Clip	165
	EMF Image	261	application/octet-stream	5
	GIF Image	13	Folder	23
	JPEG Image	626	Unknown	36
	PNG Image	221		
Glossary	.PST : Personal folders files to store all data from emails .png : Portable Network Graphics JPEG : Joint Photographic Experts Group USB : Universal Serial Bus			
	End of Report			

4.2.2.2 B – Empirical High Risk Case Scenario – e-Discovery

Case Study: TC02 Detecting unusual email communication and links

This case study was to investigate email communication and links which may have passed sensitive information or shared important business documents with people outside the company resulting in a breach of the company's policy. The company's CEO decided to investigate a suspect employee's email files and folders for further evidence. AUT Forensic Laboratory's discreet investigation confirmed that there wasn't any evidence to suggest the passing of "secret information" to undisclosed entities or external parties. However, the case findings detected some unusual email communications and links between internal employees. This case study used the Enron data-set (discussed earlier in sub-section 4.2.2.1.1), similar to the first case (TC01) scenario. The data-set, in the form of general files and folders, has few PST's archived and downloaded from the EDRM site resources.

4.2.2.2.1 Objective

The firm chose AUT Forensic Laboratory for e-Discovery services to process and produce reports on complex email messages and attachments of possible suspects.

Table 4.10: AUT Digital Forensic Laboratory Email Investigation

Action	Description
What	Digital evidence employee's email (.PST file)
Why	Test, search, process, analyse electronic files and e-mail investigation
Where	AUT Digital Forensic Laboratory
When	13 th , February, 2012.

4.2.2.2.2 Research challenge

Searching 15,568 items, 25 MB of PST electronic files, decentralised searching (full-text).

Scenario 2 preparing the email investigation for email analyses

4.2.2.2.3 Pre-processing phase

The pre-processing phase of the second test case scenario is similar to the first test case scenario discussed earlier in Section 4.1. Therefore, a similar method was used for processing electronic files, equipment and associated electronic resources.

4.2.2.2.4 Early case assessment

Early case assessment (ECA) and first-pass review are the pre-collection analysis processes and findings of the case that includes key identifiers for the specific case. It helps to develop a proportionate plan based on pre-collection analysis. During the case study, this process found some critical files that identified a relevant data source and the suspected individual's unusual communication and email links. The e-Discovery tool, *Intella 1.5.2*, was used to identify ESI content throughout the e-Discovery process. As with the first case scenario, early case assessment of case TC02 was processed through de-duplication, pre-culling, file filtering, searching and categorisation to speed up extraction.

4.2.2.2.5 Identification

A keyword search was conducted that identified a number of hits on some of the names provided. The hits included email correspondence between the subject and people with those names. Deleted emails relating to the case (sensitive information) were recovered from unallocated space. Based on the ECA report's information, the suspects (individuals) can undergo further investigation through a formal interview process. Table 4.11 shows the findings from this case study.

Table 4.11: ESI Suspects details

Suspects	Department	File Category	Message Hash
Heard	Internal	Email	4ae9f4688472347fd5c069fd43e...
Susan	Internal	Email, word document	c6104b352d03b15102969dbeca...
Paul	Internal	Email	e024d75b2e2fd9484053b0277...

4.2.2.2.6 Preservation and collection

The preservation and collection process was conducted in a similar manner to the TC01 scenario discussed earlier in section 4.2.2.1.5. In addition, the EDRM Framework guideline was used to perform systematic preservation and collection.

4.2.2.2.7 Processing and analysis

Table 4.12: Search hits for case TC02

Keyword	Count
"secret information"	0
Meet	39
Involve	0
Secret	0
Passing	6
*.jpg	2
business information	18
I am passing	4
I'm passing	4
Quotation	1
"secret Quotation"	0
capital & trade	18
trade secret	0
MD5: 4a264469a750d308369e8a7...	1

Table 4.12 shows the total search hits (count) from the suspect's PST file. Different keywords were used based on the investigation. The aim was to find whoever was passing the company's secret information to an outside organisation, therefore, the keywords were used accordingly to investigate related ESI and target individuals. Table 4.12 identifies many keyword hits and related information with the number of counts not necessarily meaning that each search hit contains secret information or documents. For example, "business information" total of hits were 18, similarly, "capital & trade" hit count was 18 but these counts were mainly related to the company's policy documents and other general documents. Therefore, an in-depth search analysis is required of the actual context search objects from the suspect's email and attachments. Figure 4.4 is an example of how the e-Discovery tool helps to find relevant links (in blue) with 'Message a Hash'.

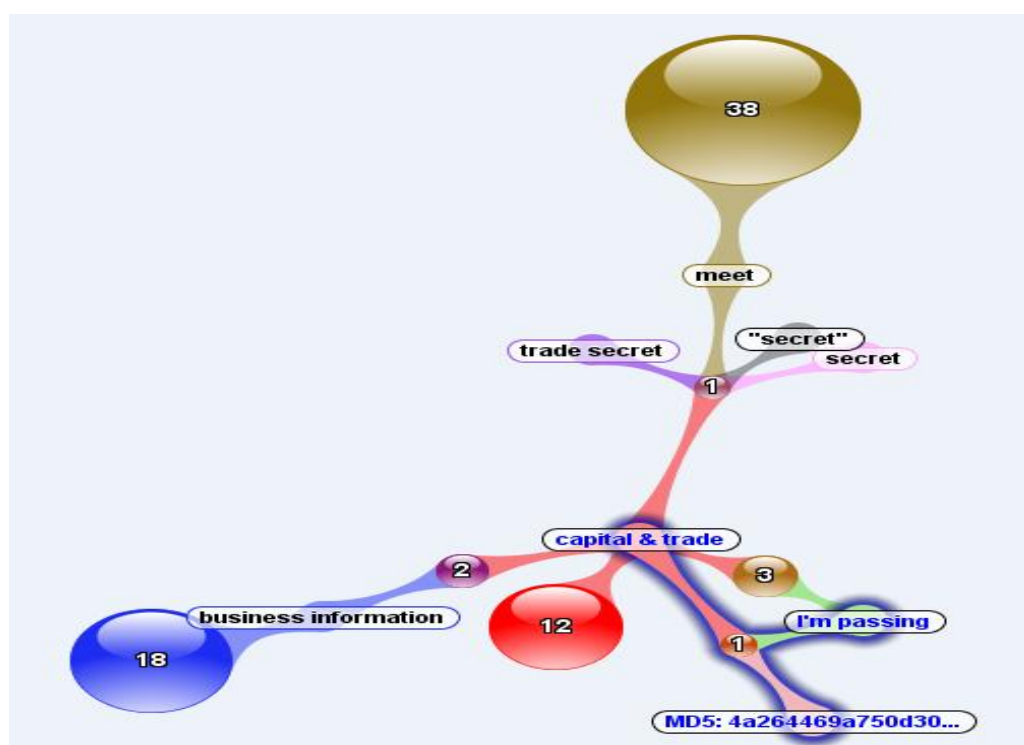


Figure 4.4: Search hits with 'Message Hash' (MD5)

Figure 4.4 is a snapshot of the keyword search hit links and the number of messages. The search methodology was enhanced by the many features of the e-Discovery tool (*Intella 1.5.2*) to confirm the findings using utilities (facets) such as; search by location, date, type, author, tags, email address, size, language and

other features. These Intella utilities help to provide in-depth analysis of any ESI. For example, *Intella 1.5.2*'s enhanced features help to identify “Hash Duplicates” keywords from an entire folder, PST file and/or any other location. Working on this case scenario, Figure 4.5 shows the “Hash Duplicates” found for specific keyword hits, helping to identify the duplicate items from the ESI.

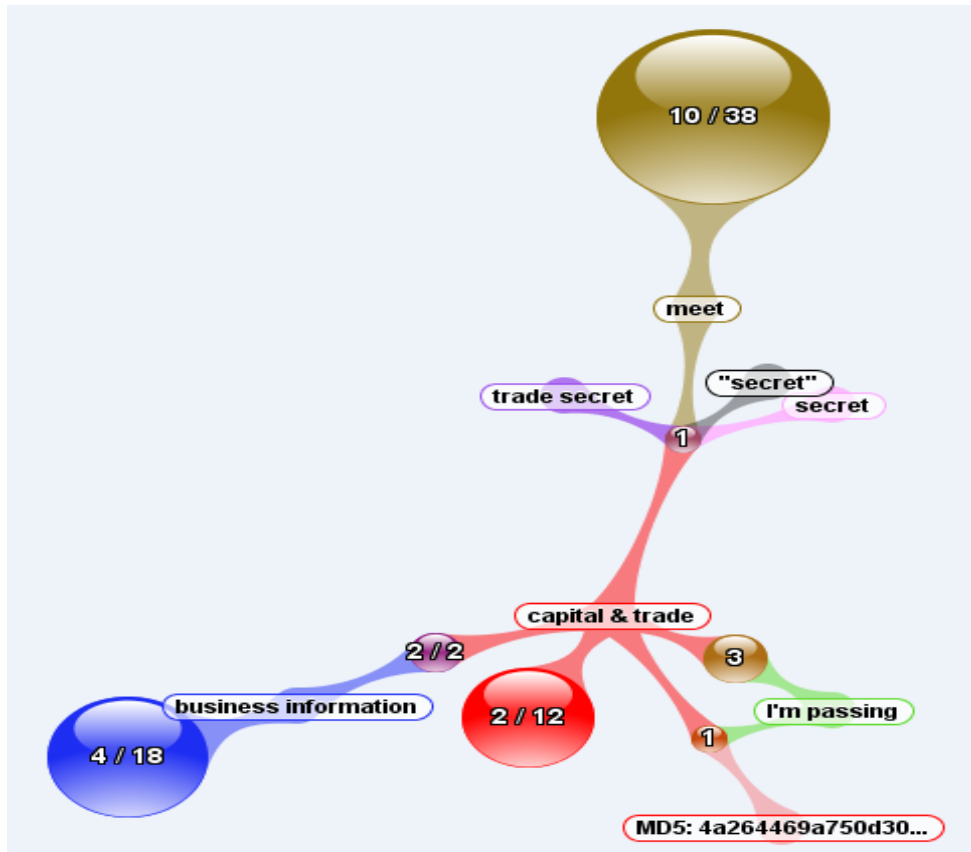


Figure 4.5: Hash Duplicates of case scenario TC02

4.2.2.2.8 Visual presentation of findings

The important step performed for this case scenario (TC02) was the recovery of “deleted” items from the suspects’ (individuals’) email and personal folder. The e-Discovery tool, *Intella 1.5.2*, has extensive features to recover deleted items from suspect list. Figure 4.6 shows the recovered items in blue from total number of search hits. These deleted items contain a lot of relevant information for case analysis.

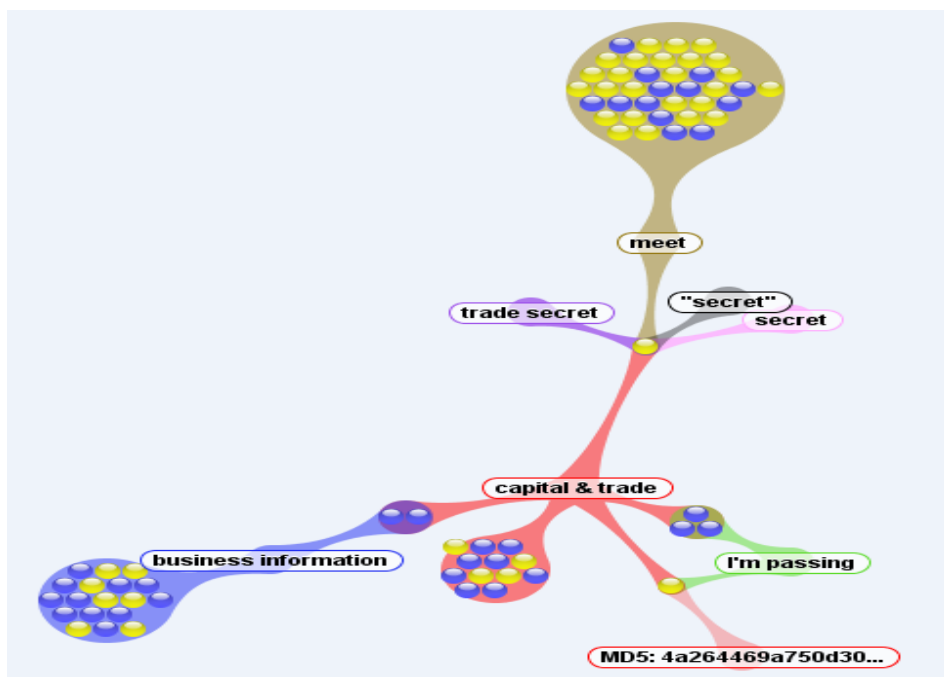


Figure 4.6: Recovered deleted items for case TC02

The methodology presented in this case scenario (TC02) combines the use of search analysis and link analysis to identify the route of internal email, links and entities. Once both analyses were constructed with the e-Discovery tool, *Intella 1.5.2*, the next step was to identify header information and message hash for integrity from the target suspect's email. Figure 4.7 shows the time stamps for the internal suspects' email communication for the given period.

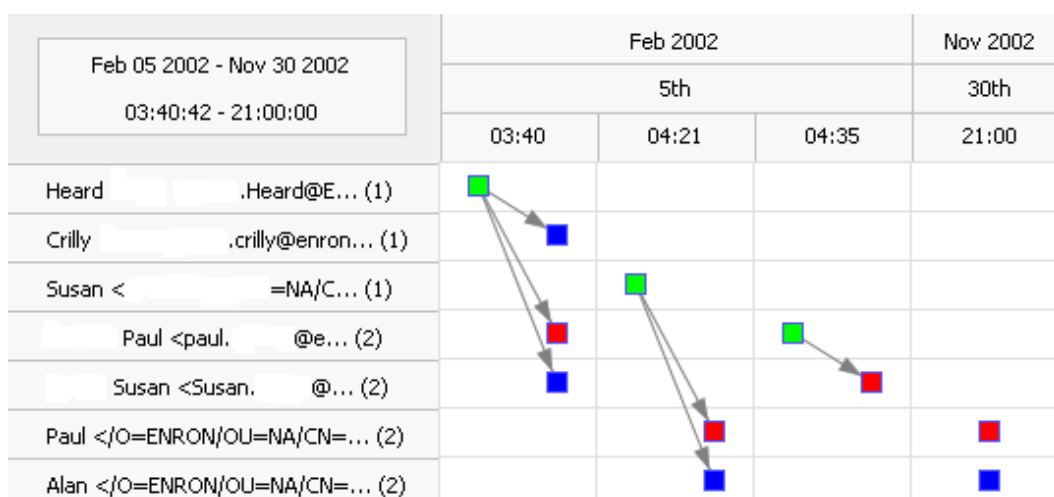


Figure 4.7: Time analysis for case TC02

In the first step, the keywords “I’m passing” found links with a total of four email messages. Once the context search found the communication between the

suspects, the next step carried over, to search by Message Hash and/or MD5 value to confirm the integrity of the communication between internal entities.

The proposed methodology recognised that, through context analysis of the individual suspect email messages, the investigation was unlikely to result in and/or prove a breach of company policy or the passing of the company's secret information to undisclosed entities. However, it must still be robust enough for the CEO and/or the attorney to make an effective decision based on sound analysis.

4.3 RESEARCH ANALYSIS

The empirical research findings for assessing the capabilities of the e-Discovery tool (*Intella 1.5.2*) are reported in this section. The section summarises the results and explains the experimental findings. In addition, the results fulfil the objectives of the main question and sub-questions of the research study. The section is divided into four sub-sections.

4.3.1 Analysis of the Empirical Low and/or High Risk Case Scenario

Results

The first test case (TC01) provided better results when compared to the second test case (TC02). Therefore, the research analysis places more weight on the first case study (TC01) to illustrate and assess the e-Discovery tool's capabilities and performance.

4.3.1.1 Test Methodology

An e-Discovery software tool, *Intella 1.5.2* was used to assess its capability. It conducted the performance testing for all case scenarios. A pilot test was carried out on a few test cases before implementing the actual case scenarios. The software performance benchmark focuses on e-Discovery processing such as; evaluating files and folders, exploring archives, extracting email attachments and other documents. The NIST file detection and duplicate file (duplicates) detection remain the same in the data-set for full chain-of-custody reporting. A search index was built based on each case scenario. The activity findings were based on private/confidential files, folders and email (.PST file) information. Once the processing steps were completed, a fully prepared report, case files and other relevant information were prepared for final output. Benchmark performance

testing was completed in a similar manner to the eDiscovery Journal Report: *Digital Reef & BlueArc eDiscovery Software Performance Benchmark Test* (eDiscovery Journal, 2010).

Table 4.13 shows the performance benchmark results for e-Discovery software (*Intella 1.5.2*).

Table 4.13: Performance analysis

Case Scenario	Data set Size	Processing Time	Processed Items			Total Items
			Files	Folders	Messages	
Case 1 (TC01)	1.0 GB	2,874 items per minute	6,958	23	8,587	15,568
Case 2 (TC02)	25 MB	1,993 items per minute	191	26	2,178	2,395

The processing speed / times noted were from the low speed (laptop) forensic work station, due to licence expiry of the software on the main forensic workstation. The processing time performance could be improved with a faster configured workstation. In general, variation in processing speed depends upon the workstation testing environment (processor, memory and other basic requirements).

4.3.1.2 Assessing the Capability of the e-Discovery Tool - Performance Analysis

The workstation 01 result was carried out on the AUT Digital Forensic Laboratory and the workstation 02 performance result was carried out on the personal laptop. The variation in processing time was due to the hardware configuration of the test machines mentioned earlier in Table 4.5. The idea of using the workstation 02 (laptop) was based on mobility, the limited licence (14 days) and to test software performance on a low speed workstation. Both experiments on the workstations successfully managed to produce results with different speed of indexing times. Table 4.14 shows the performance testing results from each test workstation.

Table 4.14: Performance testing results on test workstations

File Type	Data set Size	Indexing Time	Re-Indexing Time	Appendix Reference	Test Workstation
PST	1.0 GB	11:07 m/s	08:52 m/s	Appendix 4	Desktop (Workstation01)
PST	0.25 GB	01:12	00:52 m/s	Appendix 4	Laptop

					(Workstation02)
All files format		00:26	00:20 m/s	Appendix 4	Desktop (Workstation01)

Table 4.15 shows the results from both case scenarios specifying successful performance testing results from the e-Discovery tool (*Intella 1.5.2*).

Table 4.15: Performance testing results of each case scenario (File type search)

File Type	Case 1 (TC 01) File Count	Case 2 (TC 02) File Count
application/octet-stream	5	0
AVI Video Clip	66	0
BMP Image	1	0
Comma-separated values file	0	1
CSS style sheet	0	1
Email message	8640	2178
EMF image	261	0
Folder	23	26
GIF image	13	3
HTML document	39	5
Internet location (URL)	4427	0
JavaScript source file	0	1
JPEG image	626	1
MP3 audio clip	138	0
MPEG video clip	165	1
MS Excel document	326	42
MS Office document	340	0
MS PowerPoint document	44	1
MS Publisher document	0	1
MS Word document	54	101
MS-DOS/Windows executable	48	3
PDF document	3	0
Plain text document	23	3
PNG image	221	11
Rich text document	0	5
TIFF image	3	0
Unknown	36	1
vCard file	12	1
WAV audio clip	4	0
WMF image	50	7
XML document	0	2

4.3.1.3 Assessing the Capability of the e-Discovery Tool – Search Analysis

Keyword searching is the fundamental process of identifying any ESI and sorting out relevant information. Different types of searching techniques were performed using search methodology such as; simple keyword search, Boolean search, fuzzy search and proximate search. In some cases, conceptual searching techniques are used to identify the context of concepts in the documents that exceed keywords. Figure 4.8 represent a few keyword search hit results from the TC01 case scenario.

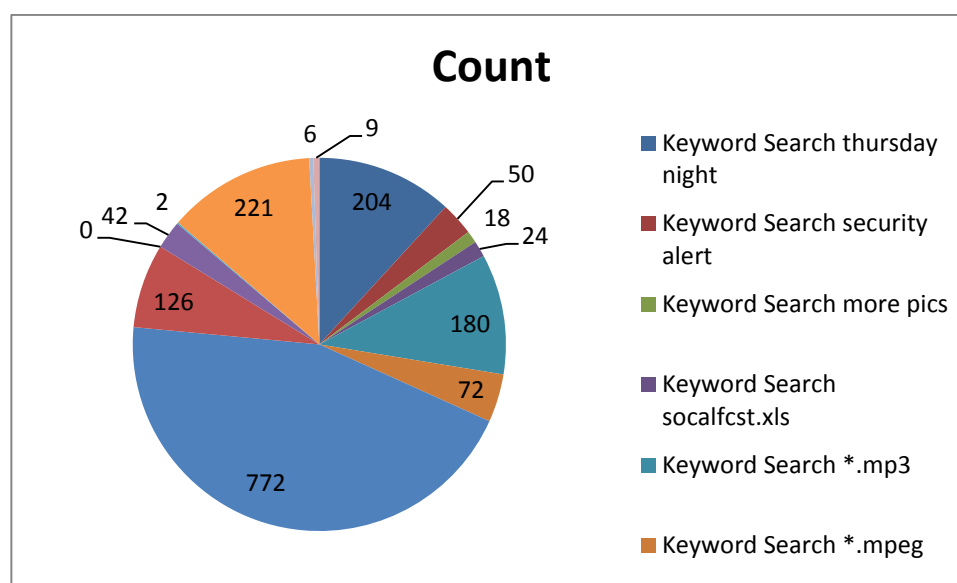


Figure 4.8: Keyword search – count in numbers

Table 4.16: Email analysis including file attachments using search methodology

Analysis of top of personal folder for suspect Zl_Lenhart:					
Total Searches (through email location)					
		Attachment - file type by extension			
Suspect (zl_lenhart) Inbox	Total Files	.png	.jpg	.mpeg	.avi
All documents	425	69	219	21	44
Contacts	0	0	0	0	0
Deleted items	67	8	8	69	126
Discussion threads	370	69	221	21	34
Drafts	0	0	0	0	0
Inbox	5	0	0	0	0
Notes inbox	11	6	2	2	0
Sent	414	63	217	19	44
Sent items	168	6	64	6	4
Total	1460	221	731	138	252

Table 4.16 shows the investigative file type analysis from case TC01 suspect's top personal folder using search methodology. The importance of these findings illustrates the search capability, simplicity and performance of the e-Discovery tool.

Table 4.17: Specific keyword search using search methodology

<u>Key word "Thursday night"</u>		
Suspect (zl_lenhart)'s inbox	Total Files	No. Of Hits
All documents	425	24
Contacts	0	0
Deleted items	67	48
Discussion threads	370	16
Drafts	0	0
Inbox	5	5
Notes inbox	11	1
Sent	414	23
Sent Items	168	87
Total	1460	204

Out of 8,626 messages a total of 204 hits were found of the related keywords "Thursday night". Table 4.17 represent the findings of the keyword search including subfolders.

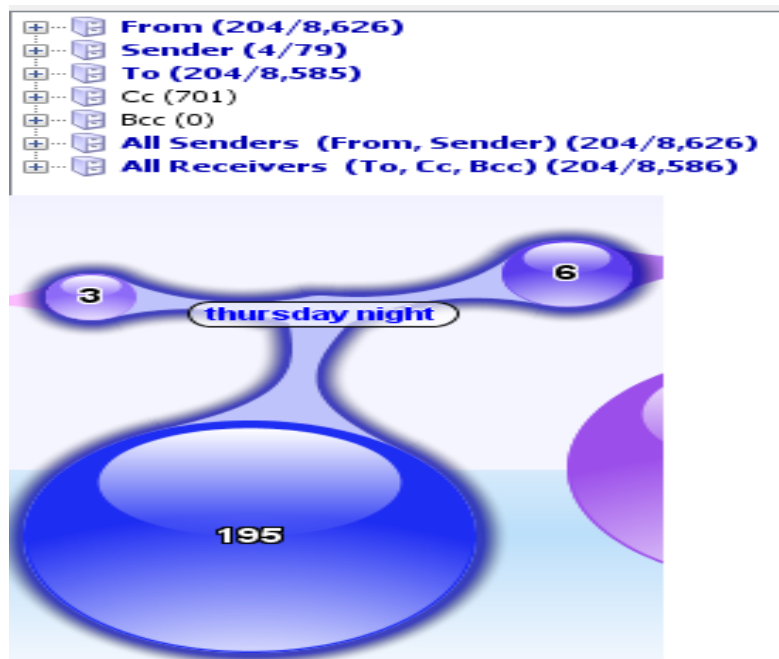


Figure 4.9: Visual evidence for "Thursday night" keyword hits.

The keywords “Thursday night” were the most significant source of evidence in case scenario TC01. All further investigation and link analysis of the suspect was carried out from this significant search hit.

Keyword searching is the foundation of search methodology. Therefore, an in-depth search analysis methodology was performed for both cases as discussed earlier in Chapter 3 (Table 3.1), to find relevant information. For example, a keyword search of ‘metadata’; contents, headers and properties of email and attached documents. Moreover, an enhanced feature of the e-Discovery tool *Intella 1.5.2* evaluates the relevant information by providing the facilities to review, explore and produce a report with print and export functionality. Therefore, the keyword searching methodology was an integral part of the empirical research findings.

4.3.2 Assessing the Capability of the e-Discovery Tool

The main analytic technique used for both case scenarios is “Clustering” which more efficiently helps to identify potentially relevant data through content relationship. In addition, the “metadata” search was used to identify which documents were edited at certain times and by whom. Using near duplicates analysis allows the system to “tag” all members within a collection of documents and/or parent emails. Section 4.3 discussed the empirical research results and analytic techniques.

The e-Discovery tool (*Intella 1.5.2*) demonstrated successful results and analysed relevant information with simplicity and efficiently. Therefore, assessing the capability of the e-Discovery tool (*Intella 1.5.2*) successfully fulfilled the objectives of this research.

4.3.3 Assessing the Capability of an open-source tool for e-Discovery

Digital Forensics Framework (DFF) is open-source investigation software. The open-source tool was verified for the same purpose to assess the capability using both case scenarios; however the execution time for processing .PST file and other data format file was reported to be more time-consuming process than expected for processing a case file. In addition, the installation, processing case files, indexing and pilot testing, performed in order to identify potentially relevant data through processing, found difficulty in processing of same case scenarios.

The raw results from open-source tool (DFF) will be discussed in following chapter 5, based on pilot testing findings, case scenarios, performance testing, analysis and verification of an open-source e-Discovery tool (See Appendix B for data).

4.4 CONCLUSION

Chapter 4 focuses on the empirical case study and the various findings produced from the process of assessing the capability of the e-Discovery tool. The main aim of Chapter 4 was to test, process, analyse and report the findings using a case scenario from a public data-set. By evaluating the e-Discovery tool and producing the results in the form of a report, Chapter 4 accomplished the objective through applying the methodology that was defined in Chapter 3.

The research analyses of two empirical case scenarios were shown in Table 4.15. Each test case scenario is illustrated with a table that summarises the findings. Both cases used the EDRM framework as a benchmark of the investigation process and various methodologies to assess the tool's capability in terms of performance, search analysis, visual presentation and reporting. The main aim of Chapter 4 is to present a framework for the analysis of unstructured and structured data by performing a practical email investigation of case scenario using the e-Discovery tool (*Intella 1.5.2*). In addition, link analysis visualisation provides an elementary analysis of the email investigation that can be used as an effective visual presentation to identify the key actors within case scenarios.

For each case scenario, two different production formats were used to display the findings. The first case scenario used a brief report format according to the NIJ brief sample report described in Sub-section 4.2.2.1.7. The second case scenario illustrates the visual representations of findings in Sub-section 4.2.2.2.8. The findings are now discussed in detail in Chapter 5 to answer the research question and sub-questions.

Chapter 5

Discussion of Findings

5.0 INTRODUCTION

Chapter 5 is a discussion of the research findings, the main research question and the hypothesis. Chapter 2 defined the importance of e-Discovery, software tools, and the EDRM framework. Chapter 3 developed the research question, the data analysis and the research methodology. Chapter 4 reported the findings based on the case scenarios, performance testing and analysis of the e-Discovery tool. The discussion of findings is to link the findings in Chapter 4 to the bigger picture raised in Chapter 2. The e-Discovery process can be further elaborated on and the research question and sub-questions answered. Different hypotheses are proposed in Chapter 3, Section 3.2.4 to answer the main research question and these will be tested in Chapter 5. The in-depth discussion will refer back to relevant sections in Chapter 4.

Chapter 5 is split into four main sections. Section 5.1 is a discussion of the research question and sub-questions outlined in Section 5.1.2. The associated hypotheses outlined in Section 3.2.4 are tested using the e-Discovery tool *Intella 1.5.2* to find answers to the hypotheses queries based on the research findings. Section 5.2 discusses the capability and importance of the e-Discovery tool. Section 5.3 discusses the limitations of this research, following on from Section 3.4. The final section, Section 5.4, discusses possible areas for further research and a conclusion.

5.1 DISCUSSION OF THE RESEARCH QUESTION

Based on the literature review in Chapter 2, and the review of similar studies in Chapter 3, the main research question and secondary questions (Section 3.2.3) were developed. The research questions will be answered based on the performance analysis completed in Chapter 4's results and report.

In this section, each table will include a question asked along with the asserted hypotheses (Section 3.2.4), from knowledge acquired from the literature review. The experimental test results will be presented as either in favour or

against the hypotheses and a judgement made accordingly.

The discussion in the table will be a judgement of the knowledge gained from the literature and the experimental research findings.

5.1.1 Main Research Question and Associated Hypotheses

The literature review in Chapter 2 highlights many key features for individual e-Discovery software. Therefore, the main research question was derived from the research methodology (Chapter 3) in order to provide a specific objective for the empirical research.

In general, companies face many challenges when it comes to choosing the e-Discovery tool that is right for them. Therefore, the aim of this research is to assess the capability of an e-Discovery tool in terms of performance, accuracy and complexity. The afore-mentioned research question was:

“What performance can be expected of e-Discovery tools when extracting evidence?”

Generally speaking, the performance of an e-Discovery tool can be assessed through its analytical capabilities including; search performance, email threading analysis, advanced visualisation and overall processing of the case and production of the report in different file formats. Therefore, the capability of an e-Discovery tool was assessed through a series of standard test cases and two major case scenarios. In Chapter 4, the capabilities and the attributes of an e-Discovery tool were evaluated through performance testing. The overall e-Discovery process from start to finish used an EDRM framework as a benchmark.

In order to answer the main research question as proposed in Section 3.2.3, a number of empirical research testing phases (such as importing electronic files, e-mails, extract text, metadata and hundreds of different types of files; de-duplication and culling) were undertaken to achieve maximum productivity for an efficient e-Discovery process.

Therefore sub-questions were formulated as follows:

- I. How quickly can the tool analyse / produce information for e-Discovery?*
- II. What is the complexity of the e-Discovery tool?*
- III. Does the e-Discovery tool obtain significant information?*

The key features and functionality of industry-leading e-Discovery tools were reviewed in Chapter 2. Therefore, testing the capability of a selected tool to answer the research question was necessary. The hypothesis of the research will be based on assessing the capability of an e-Discovery tool and how it can be achieved by using test case scenarios.

5.1.2 Secondary Research Questions and Associated Hypotheses

As stated in Section 3.2.3, a total of three secondary research questions were derived in order to support the main research question and related areas of concern. These provide in-depth analysis in order to answer the main research question.

The following tables (Table 5.1, Table 5.2 and Table 5.3) present the discussions of the answers to the sub-questions and/or secondary questions. Table 5.4 present the main research question and its summary. A statement of accepting, rejecting or considering the hypotheses will be articulated for each question based on a summary of the analysis and the research outcome of each question.

Table 5.1: Secondary Research Question 1 and Tested Hypotheses

<i>Sub Question 1: How quickly can the tool analyse / produce information for e-Discovery?</i>	
Hypothesis 1 (H1): The e-Discovery tool can be assessed by evaluating performance and speed of the results.	
Hypothesis 2 (H2): The capability of e-Discovery software and/or tools can be measured by testing and observation.	
ARGUMENT FOR:	ARGUMENT AGAINST:
The testing setup prepared for installation of the e-Discovery software tool <i>Intella 1.5.2</i> using the recommended hardware configuration identified for a ‘medium speed’ (main) forensic workstation. Hence, a ‘device	A minimum hardware configuration was identified for a ‘low speed’ laptop to test the processing speed, installation and configuration of the software’s minimum requirements. Table 4.4 shows the execution configuration

<p>manager' snapshot was used to record accurate system configuration (See Appendix 2).</p> <p>The testing environment was setup for major case scenarios to be able to calculate MD5 and message hashes, to calculate the number of duplicates, to calculate near-duplicate hashes, index archives, cache images, index content embedded in documents and mail source files for each case (See Appendix 1).</p> <p>Indexing is a vital part of a tool's performance, giving information about total number of items inspected such as; processed files, folders, messages and the processing speed of an individual case. Thus the indexing speed was recorded to evaluate performance (See Appendix 4).</p> <p>The configuration setup for major test case scenarios was used to assess performance in regard to speed of results. Hence, all searching criteria (features) were selected on the e-Discovery tool to evaluate its performance (Appendix 5).</p> <p>Benchmark performance testing was conducted for both major cases. Processing speed, data size and number</p>	<p>information for the testing environments (Chapter 4).</p> <p>In general, to achieve fast processing speeds / quickly produce information from a large dataset (>100 GB), the highest (maximum) hardware configuration is recommended. However the dataset for the empirical research was small (< 4.5 GB). Therefore, a low speed workstation was able to produce the information successfully and has been considered for the outcome of both case scenarios. However, variation in processing speed can be improved through using the highest hardware configuration.</p>
---	--

<p>of processed items were evaluated for successfully processed items. Analysis of performance is shown in Table 4.12.</p> <p>The performance testing (timing and/or speed) results were evaluated through indexing and re-indexing for the test case scenarios. Section 4.3.2 (Table 4.13) shows the performance testing results from the test workstation.</p>	
<p>SUMMARY: Significant findings, processing, analysis and outcomes from the e-Discovery tool <i>Intella 1.5.2</i> were successfully obtained from the pilot test and major case scenarios, within a short period of time. The results, after analysis, proved that the e-Discovery tool can produce quick results. Hence, the arguments made for and against proved both of the hypotheses are to be accepted by evaluated performance and successful test results.</p>	

Table 5.2: Secondary Research Question 2 and Tested Hypotheses

Sub Question 2: <i>What is the complexity of the e-Discovery tool?</i>	
<p>Hypothesis 3 (H3): e-Discovery software features and functionality have quicker processes for producing results.</p> <p>Hypothesis 4 (H4): The e-Discovery tool exhibits less complexity.</p> <p>Hypothesis 5 (H5): If an open-source e-Discovery tool is used on the same test scenarios, it will be more complex and time-consuming.</p>	
ARGUMENT FOR:	ARGUMENT AGAINST:

<p>The e-Discovery tool <i>Intella 1.5.2</i> has been analysed by assessing product features and functionality discussed earlier Section 2.5 (E-Discovery software capabilities). Therefore, the capability of the tool was assessed through the searching criteria features such as; file and documents, emails, dates, review and general search functionality (see Appendix 5).</p> <p>The tool complexity and comprehensiveness has been assessed through the keyword searching, link analysis and reporting features. (See Appendix 6 is an example of link analysis through searching criteria).</p> <p>The e-Discovery tool (<i>Intella 1.5.2</i>) provides a “facet” functionality that has been tested through the entire case investigation. The major testing queries checked to assess the capability of comprehensive features such search by tags, location, email address, date, type, author and languages. (See Appendix 3).</p> <p>Case indexing, processing, searching through the visual analysis and reporting functionality more easy to use compare to open-source e-Discovery tool (See Appendix A).</p>	<p>Section 2.5 (E-Discovery software capability) functionality and features were discussed and analysed theoretically. However, the assessing of the capabilities of each industry-leading commercial e-Discovery software tool is hampered by time and cost constraints.</p> <p>The complexity and comprehensiveness of the e-Discovery tool <i>Intella 1.5.2</i> was assessed with a few pilot cases and two main case scenarios based on Microsoft Outlook (PST) file (evidence). However, there are many other email file formats such as Microsoft Exchange Database (EDB) or MSG emails for example. Therefore, additional research requires checking the complexity of the tool and its support.</p> <p>Some minor issues were found with the main process not exiting properly while closing the application (<i>Intella 1.5.2</i>). However, this issue has been resolved in the updated version (<i>Intella 1.5.3</i>).</p> <p>While searching through the “facet” functionality of the e-Discovery tool <i>Intella 1.5.2</i>, the date format setting needs some adjustment to change the style and format of dates.</p>
--	---

<p>The key feature tested through indexing for email and hard drives, keyword searching for specific items in emails with comprehensive visual cluster-map that result in to different report format such as PDF, RTF and CSV files.</p>	<p>The case indexing, processing, searching through visual analysis and reporting functionalities were found to be more complex and time consuming using an open-source e-Discovery tool (<i>Digital Forensics Framework</i>) assessed through a few pilot cases (See Appendix B).</p>
<p>SUMMARY: The pilot testing and test case scenarios successfully processed the case results using the e-Discovery tool <i>Intella 1.5.2</i>. This proves the capabilities of the product through the successful outcome in the major test scenarios. However, empirical testing was limited in the amount of case scenario testing, due to time constraints. Therefore, more test cases with different email format testing will provide better indications of product complexity, features and functionality. In contrast, test case process speed and output proved that an open-source tool is more complex and time consuming compared to a closed-source tool. In addition, the installation, processing and result production using an open-source tool requires comprehensive knowledge of information technology as compared to a closed-source (commercial) tool. Therefore, the arguments made for and against prove the hypotheses (H3, H4, and H5) are to be accepted.</p>	

Table 5.3: Secondary Research Question 3 and Tested Hypotheses

<p>Sub Question 3: <i>Does the e-Discovery tool obtain significant information?</i></p>	
<p>Hypothesis 6 (H6): The e-Discovery tool will be able to provide relevant information from the case scenarios.</p>	
<p>ARGUMENT FOR:</p> <p>The first test case scenario (TC01) was able to produce significant information</p>	<p>ARGUMENT AGAINST:</p> <p>Significant evidence for the second test case scenario (TC02) was unsuccessful</p>

<p>successfully using the e-Discovery tool. The keyword search traced substantial amounts of evidence from the suspect's email file (See Appendix 6 for file structure analysis, link analysis and search by type).</p> <p>The second test case scenario (TC02) was also able to produce significant information (unusual email communication) using the e-Discovery tool (See Figure 4.6).</p> <p>The pilot test cases were successfully completed using the e-Discovery tool for electronic stored information (files and folders), verifying relevant information and processing. (See Appendix 4).</p> <p>The robust visual presentation feature of the e-Discovery tool <i>Intella 1.5.2</i> proved that obtaining information was not only effective for analysis and reviewing the findings, but it also successfully demonstrated the email communication links between the suspects (See Appendix 6).</p>	<p>in verifying the email communication. Therefore, additional investigation was required. However, the tool successfully assessed and processed all the information.</p> <p>The test case scenario (TC02) did not provide any significant information related to the case under investigation. This may create possibility of another suspect's email investigation who may have actually engaged in sharing important business information.</p> <p>The test cases were limited to processing .PST files, general files and folders. Therefore, additional research requires assessing the tool with all other types of supporting email files and folders.</p>
--	--

<p>The report generated from the e-Discovery tool (<i>Intella 1.5.2</i>) proved its capability in obtaining significant information for the relevant case scenarios tested. For an example, see the findings for case TC01 presented in Section 4.2.2.1.7 in chapter 4. The report findings for case TC02 proved that the tool successfully produced the relevant information. For instance, the IP address identified from the suspect's email file proved to be a private, internal IP address from the network and the communication within the company (See Appendix 7).</p>	
<p>SUMMARY: After analysing all significant evidence from case scenario TC01 and case scenario TC02, the results verified that the first case (TC01) was successfully able to produce significant information, therefore the test was accepted. The second case (TC02), was verified through in-depth analysis, even though no evidence was found from the suspect's email file. Nevertheless the e-Discovery tool was able to provide significant information for the case (TC02). Hence, the argument made for against proves that hypothesis H6 is to be accepted in this research experiment. Even though there was no evidence found in the suspect's PST file, it can still be considered a positive outcome for case scenario TC02.</p>	

Table 5.4: Main Research Question and Tested Hypotheses

<p>Main Question: <i>What performance can be expected of e-Discovery tools when extracting evidence?</i></p>	
<p>Hypothesis 1 (H1): The e-Discovery tool can be assessed by evaluating performance and speed of the results.</p> <p>Hypothesis 2 (H2): The capability of e-Discovery software or tools can be measured by testing and observation.</p> <p>Hypothesis 3 (H3): e-Discovery software features and functionality have quicker processes for producing results.</p> <p>Hypothesis 4 (H4): The e-Discovery tool exhibits less complexity.</p> <p>Hypothesis 5 (H5): If an open-source e-Discovery tool is used on the same test scenarios, it will be more complex and time-consuming.</p> <p>Hypothesis 6 (H6): The e-Discovery tool will be able to provide relevant information from the case scenarios.</p>	
<p>ARGUMENT FOR:</p> <p>Development of Test Scenarios:</p> <p>A pilot test run checked for different files formats through indexing and re-indexing, to measure performance before the actual ‘test case’ process using the e-Discovery tool (<i>Intella 1.5.2</i>) (see Appendix 4).</p> <p>Data Processing :</p> <p>Data processing is performed through adopting certain strategies for processing e-Discovery data such as;</p>	<p>ARGUMENT AGAINST:</p> <p>Data Processing:</p> <p>The e-Discovery tool <i>Intella 1.5.2</i> failed to process and/or indexes the direct compressed (zipped) folder as a source file. When new sources or new cases are selected with .PST in a zipped format, it gives an error message (see Appendix 8). Therefore, it requires the unzipping of the data and/or PST file – first and then selecting the data set with the .PST format. This was found to be a time consuming process (see Appendix 9, showing evidence of the processing</p>

<p>saving metadata (complete detail), deduplication, identifying redundancies, search strategy and culling data (Section 3.3.2).</p> <p>Volume Reduction:</p> <p>Section 4.2.2.1.6 stated how the e-Discovery tool (<i>Intella 1.5.2</i>) processed and successfully completed volume reduction from a large number of files (1460) to the relevant information (204). Figure 4.1 illustrated, how quickly the tool can reduce the volume into relevant information (See Chapter 4, Figure 4.1).</p> <p>File Encryption / Decryption :</p> <p>File encryption features automatically made it possible to view encrypted files.</p> <p>While processing the case, it was noted that the enhanced features of the e-Discovery tool (<i>Intella 1.5.2</i>) processed and opened the encrypted files automatically. The successful test was performed on a DKB-CV.doc (word file attachment with password protection).</p> <p>This will save time decrypting the file separately using another tool. For example, <i>FTK 1.81.6</i> listed the DKB-CV.doc file as an 'encrypted file', therefore requiring the <i>PRTK</i> tool to</p>	<p>time for a 1.41 GB file).</p> <p>Data Analysis:</p> <p>The sorting performance in the 'detail table' of the individual case needs improvement, in terms of group categories, senders and receivers columns.</p>
--	---

<p>decrypt the file separately.</p> <p>Data Analysis :</p> <p>As per section 3.3.3, key aspects measured to identify the capability of e-Discovery tool while doing data analysis such as indexing were identified, the hashes and duplications options were identified (see Appendix 2), the relationships in the content were identified via metadata analysis and important keyword searching was identified (See Appendix 6).</p> <p>Features / Functionality:</p> <p>The major features of the e-Discovery tool <i>Intella 1.5.2</i> were tested for tool functionality and significant output (See Appendix 11).</p> <p>Presentation / Report:</p> <p>The enhanced visual presentation feature, quickly produced the search query results and presented them in ‘Cluster map’ view. In addition, it exported the results in different formats such as .PDF, CSV, PST, RTF and HTML reports (See Table 5.6).</p>	
<p>SUMMARY: In order to perform the e-Discovery and to assess the capability of an e-Discovery software tool, the research study tested the software on empirical case scenarios and the test results were illustrated in Table 4.14. Each test case scenario was illustrated with a table that summarises the findings. The EDRM</p>	

framework was used as a benchmark during the investigation process and appropriate methodologies were used to assess the tool's capability in terms of performance, search analysis, visual presentation and reporting. Overall, both cases went through with successful outcomes and produced results. Therefore, the arguments made for and against prove that hypotheses H1, H2, H3, H4 and H5 are to be accepted and hypotheses H6 can take as being under consideration outcome.

5.2 DISCUSSION OF FINDINGS

The research findings have been reported, analysed and presented in Chapter 4. Therefore, Section 5.2 will now discuss the significance of the results related to e-Discovery and the four phases of research testing. Hence, the discussions will include each phase of research, the tested case scenarios and important findings in terms of answering the main research question and sub-questions.

5.2.1 Discussion of Research Phases Conducted

The empirical research was divided into four phases illustrated in Figure 3.8 (Section 3.2.5), each with a specific aim. Discussion of the research testing phases will be conducted in order to identify and emphasize significant findings (Chapter 4). The discussion of research testing phases will comprise; assessing the scope of the selected e-Discovery tool (Chapter 2), data collection (Section 3.3.1), data processing (Section 3.3.2), data analysis (Section 3.3.3) and data visualisation and/or presentation (Section 3.3.4).

5.2.2 Discussion of the Scope of the Selected e-Discovery Tool (Phase 1)

Phase 1 of the empirical research testing was to decide the scope of the study the e-Discovery tool. As stated in Section 2.3, from the reviewed literature, the range of e-Discovery software created many opportunities for forensic investigation and law enforcement, however only a few software applications are widely popular and accepted in court. In addition, the range of e-Discovery software provides a lot of functionality and many features. Industry-leading e-Discovery software tools licences are very expensive to buy and there are few with a limited licence version. Moreover, it would be an expensive approach to test and/or assess each

industry-leading tool. Therefore, the scope of the study was to assess the capabilities of a selected tool.

As stated in Section 2.6, an e-Discovery project can be seen from many different points of view, such as the business perspective, the legal perspective and the digital forensic investigative perspective. The expected outcome of software testing can be identified through assessing its capabilities. In addition, digital forensic experts are required to use the best e-Discovery tools that speed up the legal process incurring minimum cost and minimum time. Hence, the functionality and features of an e-Discovery tool have been assessed through pilot tests and case scenarios. From a non-technical perspective, e-discovery software should provide user-friendly features, should be easy to use and less complexity.

Most industry-leading e-Discovery tools boast of their key features, benefits and functionality (as discussed in Section 2.3). None would boast of complexity and comprehensiveness. However, these can be checked through assessing the capability of the e-Discovery tool. Without assessing the capabilities and understanding the scope of the e-Discovery software, it is difficult to choose an appropriate tool for a specific e-Discovery. Incorrect selection may delay the process and time constraints are also important. In addition, the features and functionalities of e-Discovery software tools have different approaches; hence, it is always challenging for an examiner to select the right software for an e-Discovery solution. Therefore, software needs impressive capabilities, and to be firm and fully functional as mentioned in Section 2.5.

5.2.3 Discussion of the Identification, Preservation and Data Collection (Phase 2)

Phase 2 is about data identification, preservation and collection of data. Therefore, data has been collected as per data requirements (Section 3.3). As stated in Section 3.3.1, data could be collected from many possible sources (Table 3.5), however, as stated in Section 3.3 the Enron Dataset was used to assess the capability of the e-Discovery tool. There were two main reasons for using the Enron public dataset. First, it was similar to real data or live data for real-world e-Discovery on which to assess e-Discovery tool performance. Second, the dataset is a useful source for research in fields like link analysis, social networking analysis and search analysis.

Referring back to Chapter 2; Section 2.4 introduced the EDRM model as a benchmark of e-Discovery. As per EDRM guidelines steps were followed for identification in which relevant information was located. For instance, an ESI checklist from EDRM was generated for the case scenario investigations (See Appendix 10), to identify potential sources of information. Preservation and data collection are an important part of digital investigation and/or e-Discovery as discussed in Section 2.4.3. Therefore, an anti-static plastic bag was used to preserve the evidence (USB Stick) and the original dataset was transferred through ‘UltraBlock Forensic USB Write Blocker’ to the digital forensic workstation’s hard drive for further processing (Section 4.1).

5.2.4 Discussion of the Test Case Scenarios’ Processing, Review and Data Analysis (Phase 3)

Testing the performance of and assessing the capability of the selected e-Discovery tool was conducted in Phase 3. Therefore, Phase 3 is the core part of this empirical research. As per the EDRM guideline, all collected and preserved digital information were verified and examined at this stage. Therefore, both case scenarios were processed with de-duplication, pre-culling, file filtering, searching and categorisation. Table 5.5 shows the evidence for each case processed, reviewed and analysed as per EDRM framework guidelines.

Table 5.5: Case scenarios processing, review and analysis

EDRM Stages	Case 1 (TC01)	Case 2 (TC02)
Processing	Figure 4.1 shows volume reduction.	Table 4.12 shows the search hits for case TC02.
Review	Figure 4.2 shows file type search.	Figure 4.4 shows the search hits through message hash, Figure 4.5 shows the hash duplicates for case scenario TC02.
Analysis	Figure 4.3 shows time analysis for email communication.	Figure 4.7 shows time analysis for case TC02.

The performance testing results for each case scenario is illustrated in Table 4.13 and Table 4.14 (Section 4.3.1 and 4.3.2 respectively). In Chapter 4, the capability of selected the e-Discovery tool *Intella 1.5.2* was assessed through different types of analysis, Table 4.15 shows performance testing results through file type search for both case scenarios. Similarly, Table 4.16 presented email analysis including file attachments using the search methodology.

Phase 3 successfully processed, reviewed and analysed the information following the EDRM framework. The test cases were used for processing in order to assess the capability of the e-Discovery tool and the overall e-Discovery.

5.2.5 Discussion of the Data Visualisation / Presentation (Phase 4)

As stated in Section 2.4.6, presentation is an important step in the e-Discovery and digital forensic investigation process. Therefore, from the beginning to the end of the e-Discovery investigation, the EDRM model was used as a guideline to present the case results as output. Phase 4, will discuss data visualisation / presentation and provide recommendations for future research. The e-Discovery tool *Intella 1.5.2* demonstrated successful results for both test cases by analysing relevant information with simplicity and little comprehensiveness. In addition, the enhanced visualisation feature (cluster map) successfully demonstrated the findings from search analysis, link analysis, and time-line analysis (See Appendix 6) from email communication between suspects.

Chapter 4, Section 4.2.2.1.7 presented the findings from test case TC01 in the form of a brief special report and Section 4.2.2.2.8 presented the visual presentation of the findings from test case TC02. The capability of the e-Discovery tool *Intella 1.5.2* has been assessed in terms of its producing results in various formats such as PDF, RTF and CSV. Table 5.6 shows the list of result exported and assessed via different formats.

Table 5.6: List for results exported

Result Export	Capability Assessed
Export to PDF	✓
Export to PST	✓
Export to destination folder	✓

Export template	✓
Exporting to a CSV file	✓
Export to original (native) files format	✓
Export to iBase and Analyst's Notebook	✗

The original exported reports were intentionally not presented due to privacy concerns. However, important findings were presented to demonstrate the successful outcome from the case scenario (See Appendix 7).

The comprehensive range of visualisation capabilities of the e-Discovery tool (*Intella 1.5.2*) was assessed through to identifying connections, patterns and propensities in complex data sets, except in the case of the iBase and Analyst's Notebook, which was due to unavailability and the fact that it was not a requirement for the case study.

The overall process of assessing the capability of the e-Discovery tool is illustrated in Figure 5.1. It demonstrates the data visualisation process of the e-Discovery research project.

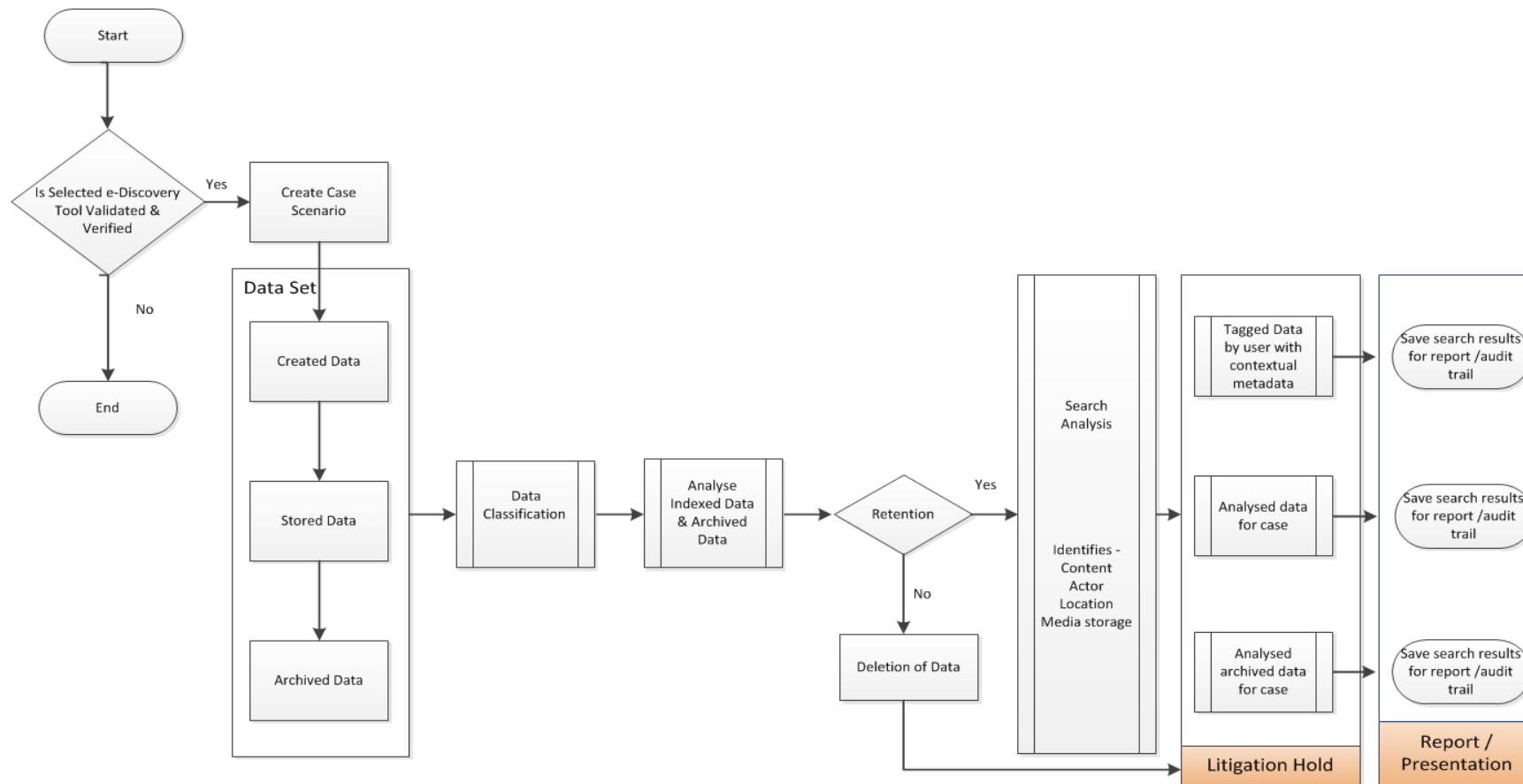


Figure 5.1: e-Discovery analysis process– Data Visualisation

5.3 DISCUSSION OF LIMITATIONS

As stated in Section 3.4, the proposed research was conducted to assess the capability of a selected e-Discovery tool (*Intella 1.5.2*). Therefore, certain limitations were expected in the proposed research. For example, the selected e-Discovery tool was only evaluated through the pilot test and two major test case scenarios. In another words, a limited set of test case scenarios were designed and tested due to time constraints and the limited licence (trial version) but with full functionality. However, all test cases ran well in terms of performance, searching capabilities, report producing functionality and other key features reviewed in Chapter 2. Hence, the proposed research successfully demonstrated the capability of the e-Discovery tool. However, there were certain limitations to assessing the full functionalities of the e-Discovery tool. For example, processing different data sets other than those proposed in this research and/or different types of ESI, such as instant messaging chats, databases and others types (voicemail, smartphones and electronic equipment). Capability assessment was limited to file formats and e-mail files (.PST) from the test case scenarios, however there are many other types of email file formats (e.g., OST, NSF, EML, DBX, IDX) and different types of ESI e-Discovery depending on the case. Therefore, a wider range of types of test cases could assess in greater depth the capability and completeness of the research.

The majority of the literature reviewed was from academic journal articles, the e-Discovery Journal and internet sources as there is, as yet, very limited research on e-Discovery software tools.

5.4 DISCUSSION OF RECOMMENDATIONS: BEST PRACTICES

The findings of the research experiment (Chapter 4) identified and executed appropriate methodology and procedures for assessing the capabilities of an e-Discovery tool and overall e-Discovery, using test case scenarios with successful outcomes. Hence, the knowledge acquired during the research experiment will now be discussed as recommendations for effective e-Discovery.

According to Osterman Research (2010), the changes to FRCP Rules 16, 26, 33, 34, 37, 45 and revisions to Form 35 are objectives for ESI. Therefore, the

changes reflect the e-Discovery process in critical ways for every litigated case. Hence, every organisation should view the body of rules from the FRCP as recommendations. In addition, the following recommendations could be taken as a main source of ideas for empirical research on e-Discovery investigation.

5.4.1 Focus on Policies for Retention and Deletion

Preserving ESI is a vital requirement for litigation. Therefore, it is important for a company to retain all of its ESI for possible e-Discovery and litigation purposes. Hence, updated policies and procedures must ensure compliance with regulatory and statutory requirements. As stated in Section 2.4.3, ESI must be destroyed under an organisation's destruction policy so that they may respond to an e-Discovery request more quickly.

“In the case of *Leon vs. IDX Systems Corporation* [2006 U.S. App. LEXIS 23820 (9th Cir. Sept. 20, 2006)], the plaintiff deleted 2,200 files from the laptop computer his employer had issued to him. The court dismissed the case and awarded the defendant \$ 65,000 for spoliation” (Osterman Research, 2010, p.9).

Therefore, it is recommended that organisations focus on policies for retention and deletion of ESI for litigation purposes.

5.4.2 Reducing the Cost of Email and Record Management

In general, reducing the cost of email and record management is a challenge for small businesses as well as a large enterprise. For instance, the challenges with current approaches using Microsoft email applications is that they can be stored in numerous locations (i.e. .PST file, Microsoft SharePoint, backup tapes and hosted on third-party solutions) across the company network. However, this can be maintained and controlled through software application. For example, Microsoft has delivered integrated email archiving, retention, and discovery capabilities with the release of Exchange Server 2010. It works efficiently to organise and manage staff collaboration and interaction. The built-in features (“Figure 1: Integrated email archiving features of Exchange Server 2010”, Microsoft, 2011, p.3) from Microsoft Exchange Whitepaper shows a strategy to preserve and discover email without having to alter either its use and/or require professional IT experience.

The figure illustrates that email can be preserved through personal archives, move and deletion policies and the hold policy. In addition, the multi-mailbox search feature streamlines the e-Discovery process.

There are many hypothetical discussions on e-Discovery blogs by experts about record management with the aim of reducing the cost of e-Discovery. However, an ideal solution is yet to be found for reducing the cost of e-Discovery and record management. For example, an automated classification of context management and an enterprise library which stores all content throughout its lifecycle to preserve ESI. However, the general recommendations for record management are reducing email size by reducing large attachment size, reducing the old content volume of ESI on primary storage and organising archival storage for litigation requirements. Therefore, an effective policy for content management will improve the speed as well as the process of an e-Discovery investigation.

5.4.3 Identification of Potential Relevant Information / Data

In general, the proactive approach helps to identify potentially relevant data for litigation requirements. The most efficient and effective approach found through this empirical research was standard search methodology to identify relevant data or relevant areas of information for the case. However, “there has been no standard method to achieve the task of searching for and reliably and efficiently locating that data” (Akers et al., 2011, p.4). Therefore, relevant tools, techniques, knowledge and experience can be used in order to meet requirements. Relevant data includes, but is not limited to the claims for ESI. Consequently, relevant data might also include particular areas of interest for litigation for example, the accounts department or administration. Once the relevant area has been identified, the next area of investigation is to identify the key players who might have possession of and/or who have created potentially relevant data. The next step is to design the questionnaire that searches basic information such as computers, data storage, backups and email accounts from the key player. For example, Table 5.7 shows potential sources of relevant information that may assist in designing a questionnaire.

Table 5.7 List of Potential Sources and Basic Information

Potential Source	Basic Information
Systems	Computer number, name, type (such as desktop, laptop, PDA), function (personal, research, business), owner and location.
Data Storage	Type (such as file server, shared drive, external drive, flash drive, DVD, CD or tape), location, ownership and property information.
Backups	Type (such as internal, external, local), ownership and location.
Mail Services	Type (such as Webmail, MSN, AOL, and Outlook), business use / personal use and email address.

Table 5.7 provides the basic information to design individual questionnaires. However, the complete form may include much more information, based on the requirements of the legislation. This information will allow IT staff and/or the investigator to more efficiently and effectively assist in data preservation requirements, in compliance with federal regulations. Once, the information is identified, a data map for the litigation should be drafted accordingly and identified, key player interviews can be considered. The individual interview provides more productive information and detailed analysis of business processes, electronic resources and the location of potential relevant data.

5.4.4 Deploy the Right e-Discovery Tool and a Proactive Approach

From an IT (technical) perspective, the case investigation and e-Discovery process was improved through early case assessment: classification of context, de-duplication, pre-culling, file filtering and searching (Section 4.2.2.2.4). For example, volume reduction of ESI and/or reduction of email size to relevant information (Section 4.1.2) will increase the processing speed of e-Discovery. However, early case assessment process is possible through an e-Discovery tool with better functionality and capability. Therefore, a proactive approach to understanding the capability of an e-Discovery tool will enhance the performance and process of e-Discovery. In addition, the e-Discovery tool must perform within

the overall processing time of the investigation and produce a successful outcome within the time frame. Moreover, the complexity of the software tool is an important factor for case investigation. For example, an open-source e-Discovery tool (*DFE*) was experimented on to process the same datasets and/or case scenarios in the proposed research without successful outcomes (see Appendix B). However, the empirical research study successfully managed to assess the capability with a closed-source (commercial) e-Discovery tool and understand the basic requirements of e-Discovery. In addition, the proactive information management strategy for 'e-Discovery cost reduction' will reduce the cost of e-Discovery. Therefore, implementing effective email archiving and retention policies are a vital way of improving the overall e-Discovery process.

In summary, businesses should protect and preserve ESI; this is essential to meet regulatory requirements for data retention and/or for legal hold. As stated in Section 2.7, to minimise the costs of conducting e-Discovery, assessing the capabilities of e-Discovery software tools as well as having a proactive approach requires implementing certain policies. In addition, effective email archiving, preservation of ESI and record management, can be an advantage in e-Discovery. Barker et al. (2008) suggested many rules regarding management policies and strategies for maintaining ESI to minimise the cost of e-Discovery and legal requirements. This research project identified and recognised rules and recommendations through the empirical experiment.

5.5 CONCLUSION

This chapter discussed the findings from the research experiment presented in Chapter 4. The answers to the proposed research questions from the Chapter 3 (methodology) were discussed in relation to the asserted hypotheses (Section 3.2.4) and a conclusion was reached with regard to the validity of the hypotheses. Likewise, the findings in regard to the e-Discovery tool's capabilities were also discussed and evaluated through the results of the test cases.

The main research question was the key source of assessment of the capability of the selected e-Discovery tool and the performance, analysis and findings. Secondary research questions and asserted hypotheses were discussed in order to support and prove the main research question, related areas of concern.

Therefore, this chapter provides in-depth analysis, while answering the main research question (Section 5.1.2). The significance of the test cases are discussed through the empirical research findings of the e-Discovery tool and demonstrated by assessing the capabilities through the testing (Chapter 4). The key recommendations are discussed in this chapter with regards to various problem areas and a summary of e-Discovery issues (Section 2.7, Chapter 2). The availability of the configuration tool for the testing environment was limited and it restricted the ability to run different types of test cases for the research discussed (Section 5.3). However, the research has made contribution to the tool evaluation research by assessing the capability, configuring tool and a comprehensive evaluation methodology. In addition, other researcher's idea, the challenges and problems are inherent in the research.

In the following chapter 6 a summary of the research conducted and the significant answers to the research questions will be summarised and areas for future research outlined.

Chapter 6

Conclusion

6.0 INTRODUCTION

Chapter 6 presents a summary of the findings and discusses a set of possible further research topics arising from this research. Overcoming the significant challenges in e-Discovery research related to assessing the capability of e-Discovery tools is an incomplete task. In Chapter 1 the research gap was identified and the remainder of the thesis has focused on assessing one particular software against constructed scenarios. The relevant literature was reviewed in Chapter 2, including literature regarding the three industry-leading e-Discovery software tools (Section 2.3), the EDRM framework (Section 2.4), the e-Discovery tools' capabilities (Section 2.5), and expected software outputs (Section 2.6). A summary was made of the relevant issues and problems (Section 2.7).

In Chapter 3, six relevant academic journal articles were reviewed and analysed to find out how such studies were conducted and which techniques were implemented in each phase of research. A research problem and question were identified (Sections 3.2.3 and 3.2.4) and a feasible research methodology specified (Section 3.3). The essential part of the e-Discovery project was the data collection strategy and preservation strategies that have been completed as per the EDRM collection guide and suggested methodology for acquiring ESI for project requirements. In the analysis phase, all relevant information for the project was identified through the fact finding, search enhancement and review enhancement strategies (Section 3.3.3). Once the major capabilities were identified in the data analysis stage, a report was generated for the defensive audit trail or as a final presentation report. The findings of the research were analysed and presented in Chapter 4. Chapter 5 discussed the main research question, the hypothesis and recommendations.

The following sections are structured to conclude this research. The research findings are summarised in Section 6.1, while Section 6.2 summarises the answers to the research questions. Future research opportunities arising are explained in Section 6.3, followed by the conclusion (Section 6.4).

6.1 SUMMARY OF FINDINGS

The findings of this research came from assessing the capabilities of e-Discovery tools, evaluation procedures and the performance of the selected tool. The main issue relating to testing procedures and execution was identified to be the availability of industry-leading e-Discovery software. This impacted on the number of tools that could be tested and the time they would be available. Therefore, the limited availability of an e-Discovery tool impacted on the scope of what could be achieved. The tool was selected to satisfy the feasibility criteria and was sufficient to demonstrate its performance and capabilities in accordance with EDRM expectations.

Table 4.13 shows the summary of performance analysis for both case scenarios. The total number of items found includes files, folders and messages (Chapter 4). In addition, Table 4.14 shows performance testing through the indexing and re-indexing times of processing each case on different workstation. The capability of the e-Discovery tool was assessed through the file type search methodology. Table 4.15 shows capability performance as per file type search for each case, including the total number of files assessed in each case and file count for each file type.

Chapter 4 identifies the major findings from the test case scenarios. Overall, both cases tested the functionality, performance analysis and report generating features of the e-Discovery tool. However, there were limitations that have been discussed in the previous chapter. Table 6.1 reviews the summary of findings and overall e-Discovery findings from both case scenarios.

Table 6.1: Summary of findings

Test CaseTC01	Test CaseTC02
A total of 24 JPEG file attachments found explicit, out of 626 JPEG images from the suspect's personal email folder.	Out of 2,178 email messages, 11 emails detected as unusual communication between the suspects.
67 deleted files were recovered from the suspect's personal (email) folder.	A total of 101 MS Word Documents found in the suspect's PST folder.

	None of them contained any secret information.
A total of 425 document files containing images in different file type extension formats (such as .JPEG, .png) were recovered.	A total of 18 search hits found for keywords; “capital & trade” and “business information”. However none of these keywords are related to the case under investigation or were misused.
Overall e-Discovery for both case scenarios.	
Test CaseTC01	Test CaseTC02
Email Message – 55 % Internet Location (URL) – 28 % JPEG Image – 4 % EMF Image – 2 % MS Word Document – 2 % MS Excel Document – 2 % MPEG Video Clip – 1 % MP3 Audio Clip – 1 % PNG Image – 1 %	Email Message – 91 % MS Word Document – 4 % MS Excel Document – 2 % Folder – 1 %

Table 6.1 shows the overall e-Discovery for both case scenarios (i.e. summary of findings) including file types searched (in percentages) from the dataset. The e-Discovery tool (*Intella 1.5.2*) demonstrated successful results for both test cases by analysing relevant information with simplicity and comprehensiveness. In addition, the enhanced visualisation feature (Cluster Map) successfully demonstrated the findings from search analysis, link analysis, and time line analysis between suspects’ email communications. The entire e-Discovery and the assessment of the capabilities of the software used the EDRM framework as a benchmark.

6.2 ANSWERS TO THE RESEARCH QUESTIONS

The purpose of the main research question was to assess the capability of an e-Discovery tool and deploy the precise steps of EDRM framework using an appropriate methodology. Thus, the findings of the research experiment (Chapter 4) have identified and examined, through correct procedures (Section 5.4.3), using a selected e-Discovery tool (Section 5.4.4), a way to achieve a significant outcome while reducing the cost of email and record management (Section 5.4.2). Factors such as cost and the time taken to perform e-Discovery will increase, if the forensic investigator does not use a pro-active approach (Section 5.4.4). Likewise, EDRM guidelines used as a benchmark to define the steps regarding; information management, identification, preservation, collection, processing, review, analysis, production and presentation of the e-Discovery process improve the quality and reduce the costs associated with e-Discovery (Section 2.4).

The research sub-questions were derived from the main research question and the answers provided are derived from the EDRM framework used as an investigation methodology designed for the digital forensic process. Hence, the research questions were answered and evaluated against the tested hypotheses by extracting evidence from the Chapter 4 findings and using a table format presented in Tables 5.1 – 5.4. In addition, Table 6.1 shows the overall statistics of e-Discovery process for both case scenarios. However, there are limitations in the research conducted (Section 3.4) that have been discussed (Section 5.3).

In brief, the main research question was the fundamental source for this project to prove performance, analysis and findings. The secondary research questions and asserted hypotheses were important in supporting and proving the main research question (Section 5.1.2).

6.3 RECOMMENDATIONS FOR FURTHER RESEARCH

The following sub-sections will discuss general recommendations for future research. Sub-section 6.3.1, discusses recommendation for testing other e-Discovery tools. The conducted research shows that the e-Discovery tool *Intella* is ideally suited for investigating email applications. However, there are many types of email applications therefore, future research for testing different types of

applications and locations are discussed in Section 6.3.2. The final sub-section (Section 6.3.3) will discuss the area of tools and future recommendations.

6.3.1 Testing of Other e-Discovery Tools

Due to the rapid evolution in technology and current e-Discovery trends in digital forensic tools having just one tool or technology is unlikely. In addition, the growing complexity of e-Discovery requirements and time constraints will inspire researchers to assess the capabilities of other e-Discovery tools. There are many distinct features for e-Discovery software described in Section 2.3, which show the capabilities of individual tools. For example, *EnCase eDiscovery* takes the proactive approach in analysis and generates the “first-pass review” feature to make a better case strategy and faster decisions. Likewise, AccessData Group introduced *AD Summation CaseVantage* with enhanced features that provide an e-Discovery solution through a standard web browser with secure internet access (Section 2.3.2).

Here, the aim of further research into other e-Discovery tools would be to find out which e-Discovery tool is the most successful under various testing scenarios. In other words, assessing the capabilities of e-Discovery tools could be compared to this research, in terms of performance, complexity and report generating.

6.3.2 Testing Different Types of Email Applications and Locations

Electronic mail (email) is the most valuable source of evidence in civil and/or litigation cases. Therefore, further testing with different types of application such as Windows mail, Eudora, Entourage (on Apple computers), webmail provider emails (e.g., Gmail, Hot Mail or Yahoo!) and the number of locations, requires further research to evaluate dynamic performance through using various functionalities of the selected tool. For example, the most popular email client under the Microsoft Windows platform was Microsoft Outlook Express (using .dbx files) was replaced by Windows Live Mail (using .emp files) to stores mail data. Another successful open-source client Mozilla Thunderbird stores data in .MSF files. Similarly, Microsoft SharePoint contains user profiles, wikis, blogs, discussion threads, contacts, calendars and image repositories. Moreover, there are a number of file extensions related to the email data file that need to be tested

in future research. In addition, according to Osterman Research (2011, p.4), “ESI consists of a large number of data types and that may be in any number of locations”. Table 3.5 outlines the wide number of sources for various types of data residing in different formats and locations. Moreover, the increasing variety of cloud-based storage repositories such as Dropbox and SugarSync will be a research challenge in the future. Therefore, when e-Discovery takes place for litigation, it is recommended that the scope of e-Discovery should be defined to speed up the process and minimise the cost.

6.3.3 Area of Tools, Future Recommendation

There are many e-Discovery tools and techniques which make the e-Discovery faster when compared to the traditional approach. In order to mitigating of the risk involved in e-Discovery, organisations should take a pro-active approach. The Deloitte Survey for e-Discovery in organisations shows that “56% have implemented e-mail management or archiving, while 32% are in the process of doing so” (Deloitte Development LLC, 2010). Likewise, “27% implemented an electronic records management program, 29% are in the process of doing so” (Deloitte Development LLC, 2010). However, according to the survey, half of the organisation don’t know how to use and/or do not use enterprise forensic collection technologies. Deloitte Development LLC (2010) suggested the best e-Discovery approach is through the consolidation of e-Discovery technology to reduce the number of brands. Another approach was based on reassessing software capabilities to respond to the requirements of litigation. In addition, the “survey results suggest five areas for companies to consider for potential improvement in eDiscovery management”, such as implementing training, improving communication between the legal department and the IT department, controlling social media, leadership commitment and vendor consolidation (Deloitte Development LLC, 2010).

Moreover, there are the recommendations for best practice discussed in Section 5.4; Chapter 5, for an effective pro-active approach to e-Discovery. Buyer’s Guide - Guidance Software suggested the “Product Evaluation Checklist”, giving overall and comparative guidelines regarding the features, functionality and value for money for e-Discovery tools. Additionally, in the recent e-discovery blog, Clearwellsystems, the author gives an overview of

Gartner's "2012 Magic Quadrant for e-Discovery Software", an annual report that provides a useful roadmap for legal technologies and a detailed evaluation of each different vendor (Gonsowski, 2012). This e-discovery blog indirectly addresses the capabilities of e-Discovery tools that identify similar aims to this research. The motivation of that report was to assess the functionality of e-Discovery tools (i.e. an e-Discovery tool does what it says it will and will process the relevant information within the time frame claimed). Therefore, the end goal should be maintained during e-Discovery with minimal cost and minimal overall processing time, while meeting qualitative litigation requirements.

6.4 CONCLUSION

The main aim of this research was achieved. This comprised of assessing the capability of an e-Discovery tool in terms of performance, complexity and the ability to find significant information using forensically-sound methodology and the EDRM framework. Therefore, the fundamental research question was derived to check the competency of the e-Discovery tool. Secondary research questions and asserted hypotheses were discussed in order to support and prove the main research question and related areas of concern. Key recommendations are discussed in this chapter with regards to future research. The major finding of the research has shown the capability of the e-Discovery tool. In addition, the results confirm that the evidence generated by the tool can survive the scrutiny of the courts.

To conclude, the research conducted demonstrates the development of knowledge, methodology and processing steps in e-Discovery, based on tested case scenarios. The information contained in this thesis will provide best practices (Section 5.4) for digital forensic investigators, lawyers and business people that meet the overall expected software outcomes (Section 2.6). In addition, this research enriches the body of knowledge regarding the testing of e-Discovery tools by building a standard methodology using the EDRM framework. The information provided in this research could prove useful for other researchers who wish to conceive a methodology, process of e-Discovery findings using EDRM framework and understand the capability of an e-Discovery tool. However, this research has limitations and these may be opportunities for future research.

REFERENCES

- 5 Tips for Choosing eDiscovery Software. (2010, November 18). *Articlebase*. Retrieved from <http://www.articlesbase.com/>
- AccessData Group. (2011). *CASE STUDY: Top 5 Communications Company Evaluates Leading eDiscovery Solutions* [White paper]. Retrieved from <https://www.apersee.com/assets/download/77/example.pdf>
- AccessData. (n.d.). Retrieved from <http://accessdata.com/>
- a) *Forensic Toolkit 3*. (n.d.). Retrieved March 2, 2011, from <http://accessdata.com/products/forensic-investigation/ftk>
- b) *AD eDiscovery*. (n.d.). Retrieved March 6, 2011, from <http://accessdata.com/products/ediscovery-litigation-support/ad-ediscovery#features>
- Adam. (2011). EDRM Model, the foundation of the eDiscovery Process. Retrieved May 5, 2011, from <http://www.ediscovery-news.com/edrm-model-the-foundation-of-the-ediscovery-process/>
- Akers, S., Mason, J., & Mansmann, P. (2011). *An Intelligent Approach to E-discovery*. Retrieved from <http://www.umiacs.umd.edu/~oard/desi4/papers/akers.pdf>
- ArbsGVSU. (2009, April 24). What Every Businessperson Should Know About E-Discovery[Video file]. Available from http://www.youtube.com/watch?v=eo03DWk4_IU
- Arkfeld, M. (2006). *A Call for Collaborative Action*. Retrieved from http://www.lawpartnerpublishing.com/newsletter/Feb_2006_EDE_Newsletter.pdf
- Arkfeld, M. (2008). Arkfeld's Best Practices Guide for Pretrial Discovery: Strategy and Tactics. Retrieved April 12, 2011, from http://www.elawexchange.com/index.php?option=com_content&view=article&id=137&Itemid=494
- Ball, C. (2010). *E-Discovery: Right from the Start*. Retrieved from http://www.craigball.com/Ball_Right from the Start_20081106.pdf
- Barker, R., Cobb, A., & Karcher, J. (2008). The legal implications of electronic

- document retention: Changing the rules. *Business Horizons*, 52(2), 177-186. [doi:10.1016/j.bushor.2008.10.006](https://doi.org/10.1016/j.bushor.2008.10.006)
- Baron, J., & Thompson, P. (2007). The Search Problem Posed By Large Heterogeneous Data Sets in Litigation: Possible Future Approaches to Research. *ACM Digital Library*, 4 (8), 141-147. Retrieved from Expanded Academic Database.
- Bayuk, J. (2010). *CyberForensics: Understanding Information Security Investigations*. Retrieved from <http://books.google.co.nz/>
- Beckett, J., & Slay, J. (2007). Digital Forensics: Validation and Verification in a Dynamic Work Environment. *IEEE Conference Publication*. 1-10. doi: 10.1109/HICSS.2007.175
- Blumenschein, J. (2011, August 30). E-Discovery: Collection Best Practices [Video file]. Available from <http://www.guidancesoftware.com/WebinarVideo.aspx?wid=1000017172>
- Burgess, S. (n.d.). Computer Forensic, Data Recovery and E-Discovery Differ. Retrieved February 17, 2011, from <http://ezinearticles.com/?Computer-Forensics,-Data-Recovery-and-E-Discovery-Differ&id=1184255>
- Carlton, G., & Worthley, R. (2009). An evaluation of agreement and conflict among computer forensics experts. *Proceeding of the 42nd Hawaii International Conference on System Science*. 1-9.
- Chisholm, C. (2010). Integrating Forensic Investigation Methodology into eDiscovery. Retrieved April 12, 2011, from http://www.sans.org/reading_room/whitepapers/incident/integrating-forensic-investigation-methodology-ediscovery_33448
- Cisco Systems, Inc. (2001). *A Beginner's Guide to Network Security*. Retrieved from http://www.cisco.com/warp/public/cc/so/neso/sqso/beggu_pl.pdf
- Civil Discovery & Privilege Law. (n.d.). Retrieved May 10, 2011 from http://california-discovery-law.com/electronic_data_discovery_new_developments.html
- Clearwell Systems. (n.d.). Introduction the Federal Rules of Civil Procedure (FRCP). Retrieved October, 12, 2011 from <http://www.clearwellsystems.com/e-discovery-101/frcp-basics.php>
- Deloitte Development LLC. (2010). *E-discovery Mitigating risk through better*

- communication*. Retrieved from http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/FAS_ForensicCenter_us_fas-us_dfc/us_dfc/us_dfc_e_discovery_survey_final_061710.pdf
- Digital Forensics Framework. (n.d.). *Open Source Digital investigation software (DFF)*. Retrieved February 04, 2012, from <http://www.digital-forensic.org/>
- eDiscoveryJournal. (2010). *eDiscovery Journal Report: Digital Reef & BlueArceDiscovery Software Performance Benchmark Test*. Retrieved from <http://www.bluearc.com/bluearc-resources/downloads/white-papers/BlueArc-WP-Digital-Reef-and-BlueArc-Software.pdf>
- EDRM. (n.d.). The Electronic Discovery Reference Model. Retrieved March 2, 2011, from <http://edrm.net/>
- a) *EDRM Stages*. Retrieved from March 2, 2011, from <http://www.edrm.net/resources/edrm-stages-explained>
- b) *Collection Guide*. Retrieved from March 5, 2011, from <http://www.edrm.net/resources/guides/edrm-framework-guides/collection>
- c) *Analysis Guide*. Retrieved from April 10, 2011, from <http://www.edrm.net/resources/guides/edrm-framework-guides/analysis>
- d) *Enron Email Data*. (2011). *EDRM Enron Email Data Set v2* [Data file]. Retrieved from <http://www.edrm.net/resources/data-sets/edrm-enron-email-data-set-v2>
- educe.edu. (2011). E-Discovery Guideline and Toolkit. Retrieved October, 12, 2011 from <http://www.educe.edu/wiki/E-Discovery+Guideline+and+Toolkit>
- Favro, P. (2012, November 1). New Gartner Report Spotlights Significance of Email Archiving for Defensible deletion [Web log post]. Retrieved from <http://www.clearwellsystems.com>
- FRCP. (2006). *Federal Rules of Civil Procedure*. Retrieved from http://bulk.resource.org/gpo.gov/prints/109/h_31308.pdf
- Forte, D., & Power, R. (2006). Electronic discovery: digital forensics and beyond. *Computer Fraud & Security*. 2006(4), 8-10. Retrieved from <http://www.sciencedirect.com>
- Forensic Computer*. (n.d.). Retrieved March 28, 2011 from <http://www.forensiccomputers.com/forensic-search/vound-intella-1.html>

- Fulbright & Jaworski L.L.P. (2009). *E-Discovery Trends*. Retrieved from <http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Duke%20Materials/Library/Fulbright%27s%20E-Discovery%20Trends.pdf>
- Gartner. (2007). Key Issues for Electronic Discovery. Retrieved March 1, 2011, from http://www.gartner.com/DisplayDocument?id=502285&ref=%27g_fromdoc%27
- Gonsowski, D. (2012, May 29). Gartner's "2012 Magic Quadrant for E-Discovery Software" Provides a Useful Roadmap for legal Technologies [Web log post]. Retrieved from <http://www.clearwellsystems.com/>
- Gould, R. (2008). eDiscovery Perspective. Retrieved November 28, 2010, from <http://www.slideshare.net/rjgould2/ediscovery-perspective>
- Guidance Software Inc. (2010). Choosing the Right In-house Electronic Discovery Solution for our organisation [White paper]. Retrieved April 11, 2011, from http://searchsecurity.bitpipe.com/detail/RES/1295458429_475.html
- Guidance Software Inc. (2012). *EnCase® Cybersecurity*. Retrieved from http://media.govtech.net/Digital_Communities/Guidance_Software/EnCase_Cybersecurity_Government.pdf
- Guidance Software (n.d.). EnCase eDiscovery. Retrieved March 30, 2011, from <http://www.guidancesoftware.com/ediscovery.htm?cmpid=701A0000000LMA4-003>
- Guo, Y., Slay, J., & Beckett, J. (2009). Validation and verification of computer forensic software tools – Searching Function. *Digital Investigation*, 6, S12-S22.
- Haggerty, J., Karran, A., Lamb, D., & Taylor, M. (2011). A Framework for the Forensic Investigation of Unstructured Email Relationship Data. *International Journal of Digital Crime and Forensics*, 3(3), 1-18. doi: 10.4018/jdcf.2011070101
- Haggerty, J., Taylor, M., & Gresty, D. (2008). Determining Culpability in Investigations of Malicious E-Mail Dissemination within the Organisation. *IEEE*. 12-20. doi:10.1109/WDFIA.2008.8
- Hall, A. (2008, November 15). How to Buy Electronic Discovery Software.

- Retrieved March, 8, 2011, from <http://electronicdiscovery.info/electronic-discovery-software/>
- Hewlett-Packard. (2011). *Ten questions and answers about digital data disclosure in an increasingly litigious and regulated world* [White Paper]. Retrieved from <http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA1-9206ENA.pdf>
- Heikkila, F. M. (2008). E-Discovery: Identifying and Mitigating Security Risks during Litigation. *IT Professional*, 10(4), 20-25. doi: 10.1109/MITP.2008.67
- Hewlett-Packard. (2010). *Preparing for e-discovery: Your road map to compliance* [White Paper]. Retrieved from <http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA2-0561ENW.pdf>
- Jeong, R.S.C. (2006). FORZA – Digital forensics investigation framework that incorporate legal issues. *Digital Investigation*, 3, 29-36. doi:10.1016/j.diin.2006.06.004. Retrieved from www.sciencedirect.com
- Kim, J.Y., Lee, T.R., Goo, B.M., & Shin, S.U. (2011). E-Discovery support tool design and implementation of the AGENT module. *IEEE*, 538-541.
- Knox, C., & Dawson, S. (n.d.). *ISO 9001: A Foundation for E-Discovery*. Retrieved from www.umiacs.umd.edu/~oard/desi4/papers/knox.pdf
- KPMG. (2010). *Electronic Discovery Management*. Retrieved from www.acfesa.co.za/event_files/Electronic%20Discovery%20Management.pdf
- Kroll Ontrack, Inc. (2007, November 11). One Year Later: The Most Significant Electronic Discovery Cases under The New Federal Rules of Civil Procedure. Retrieved from <http://articles.technology.findlaw.com/>
- Lee, E. (2011). *AccessData to Showcase eDiscovery Software Platform and Summation Case Vantage at LegalTech New York. PR Log – Global Press Release Distribution*. Retrieved from <http://www.prlog.org/11259245-accessdata-to-showcase-ediscovery-software-platform-and-summation-casevantage-at-legaltech-new-york.pdf>
- Lee, T. R., Goo, B.M., Kim, H., & Shin, S. U. (2011). Efficient e-Discovery Process Utilizing Combination Method of Machine Learning Algorithms. *Computational Intelligence and Security (CIS), 2011 Seventh International Conference*. 1109-1113. doi: 10.1109/CIS.2011.246

- Leehealey, T. (2011). What Do Wikileaks and E-Discovery Have in Common?. *eDiscovery Insight*. Retrieved April 16, 2011, from <http://ediscoveryinsight.com/2011/03/what-do-wikileaks-and-e-discovery-have-in-common>
- Losey, R. (2009). *Introduction to e-discovery: new cases, ideas, and techniques*. Retrieved from <http://books.google.co.nz/>
- Markoff, J. (2011, March 4). Armies of Expensive Lawyers, Replaced by Cheaper Software. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Mathias, T. (2007, November 26). Data Retention Gets a Second Look. *Computerworld*, 41(48), 44-44. Retrieved from <http://search.proquest.com>
- Matthews, D.R. (2010). eDiscovery versus Computer Forensics. *Information Security Journal: A Global Perspective*, 19(3), 118-123. doi: 10.1080/19393550903074135
- Microsoft Exchange Whitepapers. (2011). Addressing E-mail Archiving and Discovery with Microsoft Exchange Server 2010 [White paper]. Retrieved April 12, 2012, from <http://www.microsoft.com/exchange/en-us/whitepapers.aspx>
- Murphy, B. (2012). Are The Software Giants Dominating The eDiscovery Market?. Retrieved August 02, 2012, from <http://ediscoveryjournal.com/2012/07/are-the-software-giants-dominating-the-ediscovery-market/>
- Netsecurity Corporation. (2008). *Netsecurity forensic labs*. Retrieved from <http://www.netsecurity.com/marketing/NetSecurity-ForensicLabs-Brochure.pdf>
- NIJ. (2004). *Forensic Examination of Digital Evidence: A Guide for Law Enforcement*. Retrieved from www.ncjrs.gov/pdffiles1/nij/199408.pdf
- NIST. (2011a). Computer Forensics Tool Testing Program. Retrieved April 16, 2011, from <http://www.cfft.nist.gov/>
- NIST. (2011b). General Test Methodology for Computer Forensic Tools. Retrieved May 15, 2011, from <http://www.cfft.nist.gov/Test%20Methodology%207.doc>
- Osterman Research Inc. (2010). *The Concise Guide to E-Discovery* [White

- Paper]. Retrieved from
<http://www.ostermanresearch.com/whitepapers/download58.htm>
- Osterman Research Inc. (2011). *Key E-Discovery Issues to Consider in 2011* [White Paper]. Retrieved from
www.ostermanresearch.com/whitepapers/download138.htm
- Pasadena, C. (2010, January 19). Guidance Software Announces First Fully Integrated eDiscovery Solution. *Business Wire*. Retrieved from
<http://investors.guidancesoftware.com/releasedetail.cfm?releaseid=438471>
- Pateriya, P.K., Mishra, S., & Samaddar, S.G. (2011). *Advances in Networks and Communications*. Retrieved from <http://books.google.co.nz/>
- Project Management and Consulting for e-Discovery and Litigation support. (2010). Retrieved November 28, 2011, from <http://e-discoverypm.com/Ourthinking.aspx>
- Qureshi, A. (2009). 802.11 Network Forensic Analysis. Retrieved March 26, 2011, from
http://www.sans.org/reading_room/whitepapers/wireless/80211-network-forensic-analysis_33023
- Richard, G.G., Roussev, V., & Marziale, L. (2007). Forensic discovery auditing of digital evidence containers. *Digital Investigation*, 4(2), 88-97. Retrieved from www.sciencedirect.com
- Sans.org. (2010). Catching Hackers On The Wire. Retrieved November 27, 2010, from <http://www.sans.org/security-training/network-forensics-1227-mid>
- Schuler, K. (2008). *E-discovery : Creating and Managing an Enterprisewide Program*. Retrieved from <http://my.safaribooksonline.com/book/-/9781597492966>
- Shetty, J., & Adibi, J. (2004). *The Enron Email Dataset Database Schema and Brief Statistical Report*. Retrieved from
http://www.isi.edu/~adibi/Enron/Enron_Dataset_Report.pdf
- Shiekman, L.Z. & Robbins, N.S. (2008). *The 2006 F.R.C.P. E-discovery Amendments: A Look One Year Later*. Retrieved from
http://www.pepperlaw.com/pdfs/E-Discovery_Shiekman01052008.pdf
- Sommer, P. (2010). Forensic Science standards in fast-changing environments. *Science and Justice*, 50, 12-17. doi:10.1016/j.scijus.2009.11.006

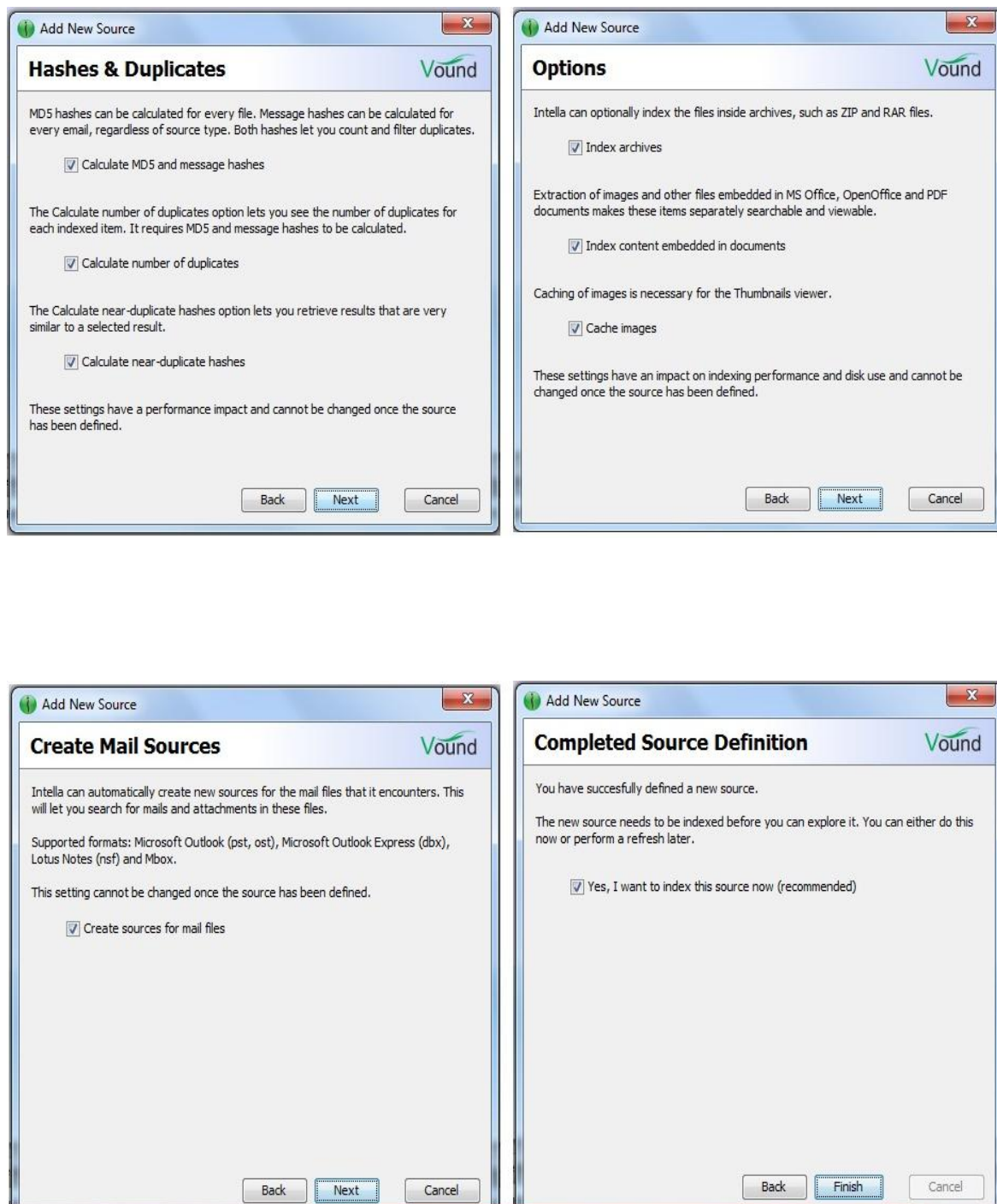
- Volonino, L., & Redpath, I. (2009a). *E-Discovery for Dummies*. Retrieved from <http://books.google.co.nz>
- Volonino, L., & Redpath, I. (2009b). *Knowing Why e-Discovery Is a Burning Issue*. Retrieved from http://media.wiley.com/product_data/excerpt/29/04705101/0470510129.pdf
- Vound Software Inc. (2012). *IntellTM User Manual*. Retrieved from <http://www.vound-software.com/docs/1.6.2/Intella-1.6.2-UM.pdf>
- Vound Software Inc. (2011). *Intella, Getting Started Guide*. Retrieved March 18, 2011, from <http://www.vound-software.com/resources/files/Intella-1.5-Getting-Started-Guide.pdf>
- Vound Software Inc. (n.d.). *Intella Product Features Comparision*. Retrieved from <http://www.vound-software.com/resources/files/Intella%20Comp%20Chart-v5.pdf>
- Wiles, J. (2007). *Techno security's guide to e-discovery and digital forensics*. Burlington, MA: Elsevier Inc.
- Willoughby, D., Hunter, R., & Antine, G. (2010). *Sanctions for E-Discovery Violations: By The Numbers*. Retrieved April 04, 2011, from <http://legalworkshop.org/2010/11/15/sanctions-for-e-discovery-violations-by-the-numbers>

Appendix A

(e-Discovery Tool – *Intella 1.5.2* (Vound Software))

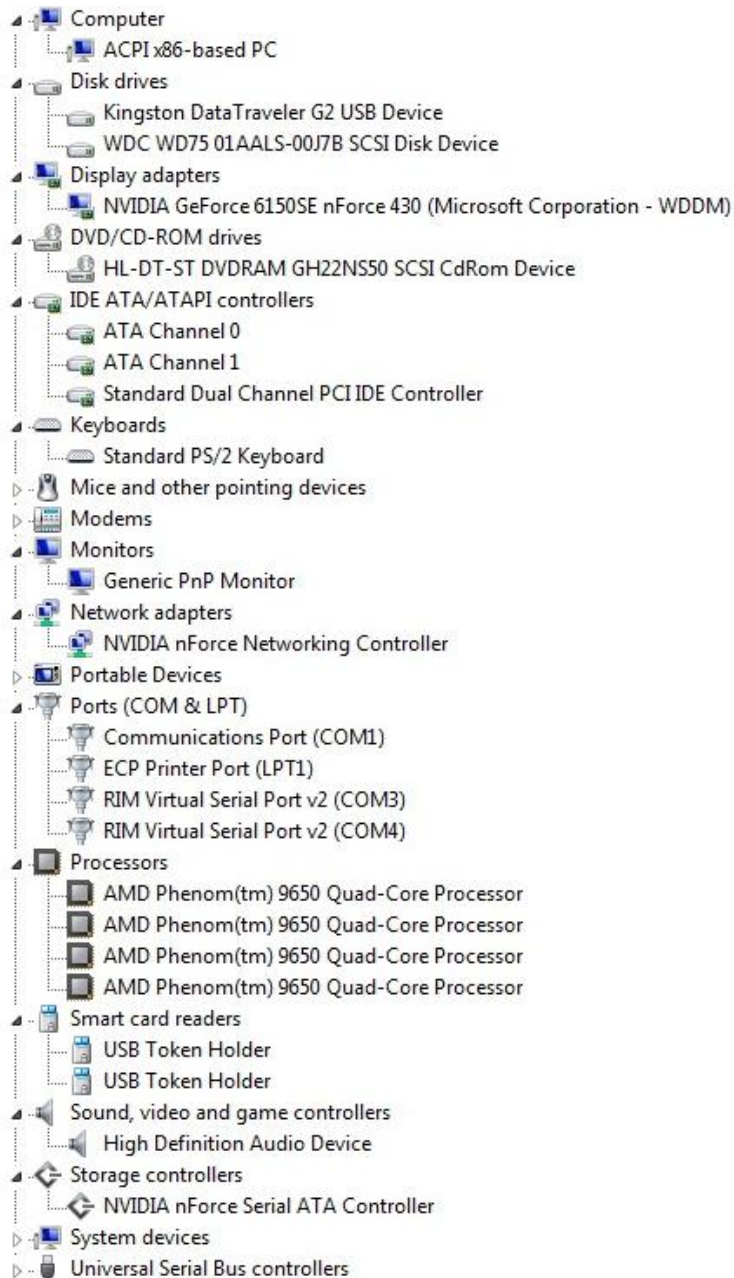
Appendix 1

Intella 1.5.2 (New Case Processing Options)



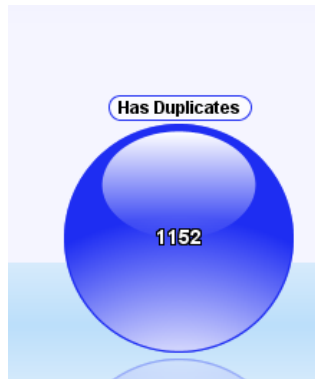
Appendix 2

Hardware Configuration (Reference: Table 4.4) (e-Discovery Forensic Workstation01 (core))

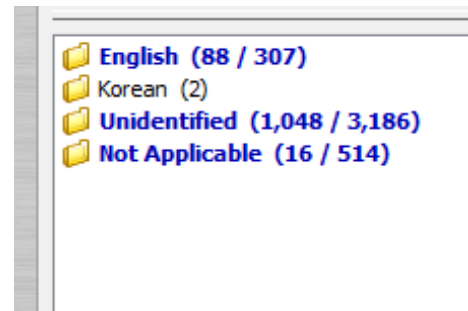


Appendix 3

e-Discovery tool Functionality, Feature - Facet



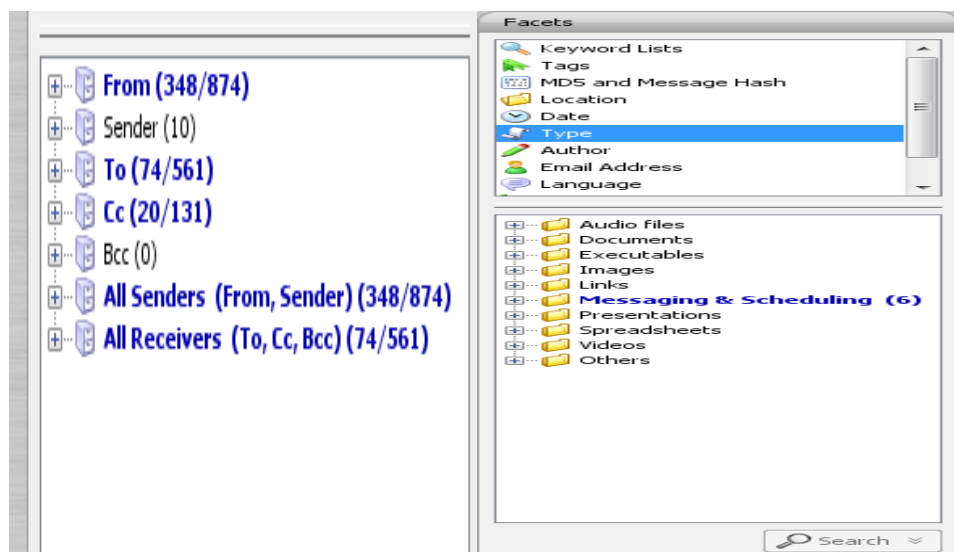
Hash Duplicates



Languages analysis



Search by Email Address:

Facets:



Appendix 4


Test Analysis & Findings [Indexing]

 **Indexing "Files Format"**  00:18

Source 1 of 1: E:\EDRM-Data Sets\Files Format

Inspected 1,401 items (all new)

Processing: tar:zip:file:/E:/EDRM-Data Sets...TAR!/force/chunker/to_swap/file_num.997

 **Indexing completed** 05:20

Processed 1 source

Inspected 14,970 items (all new)

Processed 4,829 files, 66 folders, 10,075 messages


Processing speed: 2,799 items per minute

 **Indexing completed** 00:26

Processed 1 source

Inspected 1,652 items (all new)

Processing speed: 3,675 items per minute

 **Reindex complete** 00:20

Processed 1 source

Inspected 1,652 items (all new)

Processing speed: 4,797 items per minute



Indexing completed

01:12

Processed 1 source

Inspected 2,395 items (all new)

Processed 191 files, 26 folders, 2,178 messages

Processing speed: 1,993 items per minute

Stop

Finish

Appendix 5

Configuration Process

Searching criteria / Features:

6 results

Select the columns to show in the table:

General	Dates	Files & Documents	Emails	Review
<input checked="" type="checkbox"/> Location	<input checked="" type="checkbox"/> Content Last Modified	<input checked="" type="checkbox"/> Empty Document	<input checked="" type="checkbox"/> Receiver	<input checked="" type="checkbox"/> Flagged
<input checked="" type="checkbox"/> Source Path	<input checked="" type="checkbox"/> Content Created	<input checked="" type="checkbox"/> MD5 Hash	<input checked="" type="checkbox"/> Sender	<input checked="" type="checkbox"/> Tags
<input checked="" type="checkbox"/> Name	<input checked="" type="checkbox"/> File Last Modified	<input checked="" type="checkbox"/> Creator	<input checked="" type="checkbox"/> Unread	<input checked="" type="checkbox"/> Comments
<input checked="" type="checkbox"/> Title	<input checked="" type="checkbox"/> Sent	<input checked="" type="checkbox"/> Contributor	<input checked="" type="checkbox"/> Has Attachments	<input checked="" type="checkbox"/> Previewed
<input checked="" type="checkbox"/> Subject	<input checked="" type="checkbox"/> Received		<input checked="" type="checkbox"/> Attachments	<input checked="" type="checkbox"/> Exported
<input checked="" type="checkbox"/> Type			<input checked="" type="checkbox"/> Message ID	<input checked="" type="checkbox"/> Opened
<input checked="" type="checkbox"/> MIME Type			<input checked="" type="checkbox"/> Message Hash	
<input checked="" type="checkbox"/> Size				
<input checked="" type="checkbox"/> Source				
<input checked="" type="checkbox"/> Language				
<input checked="" type="checkbox"/> URI				
<input checked="" type="checkbox"/> Duplicates				
<input checked="" type="checkbox"/> Encrypted				

For email senders and receivers show

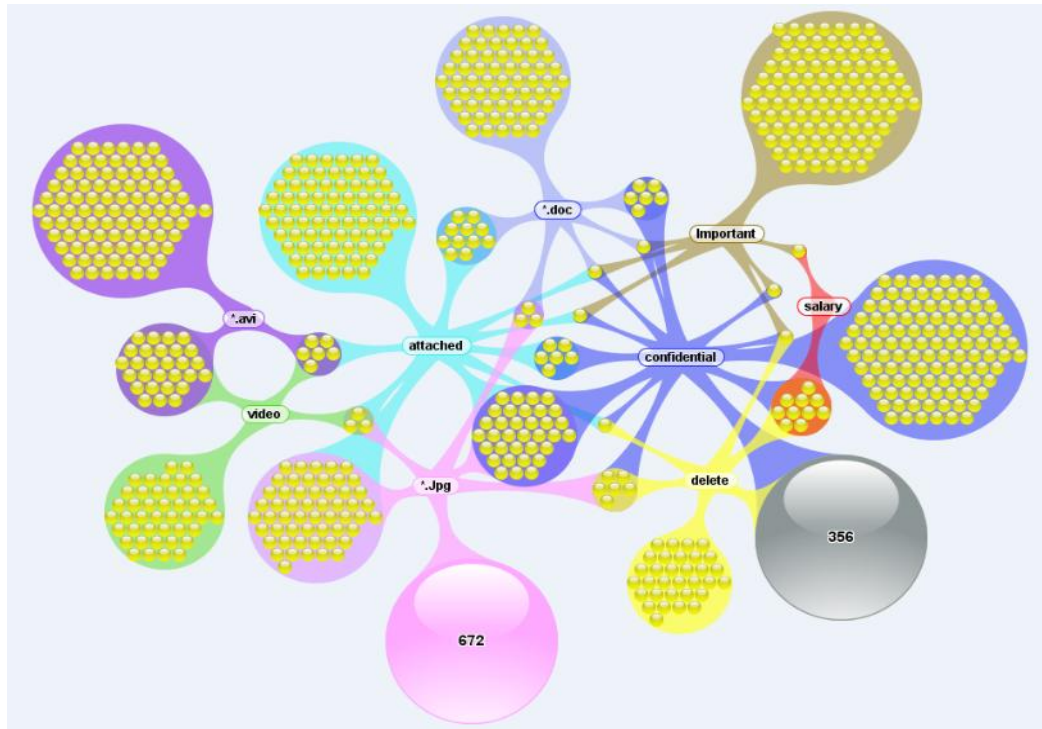
☒ Check / uncheck all

Cancel Close

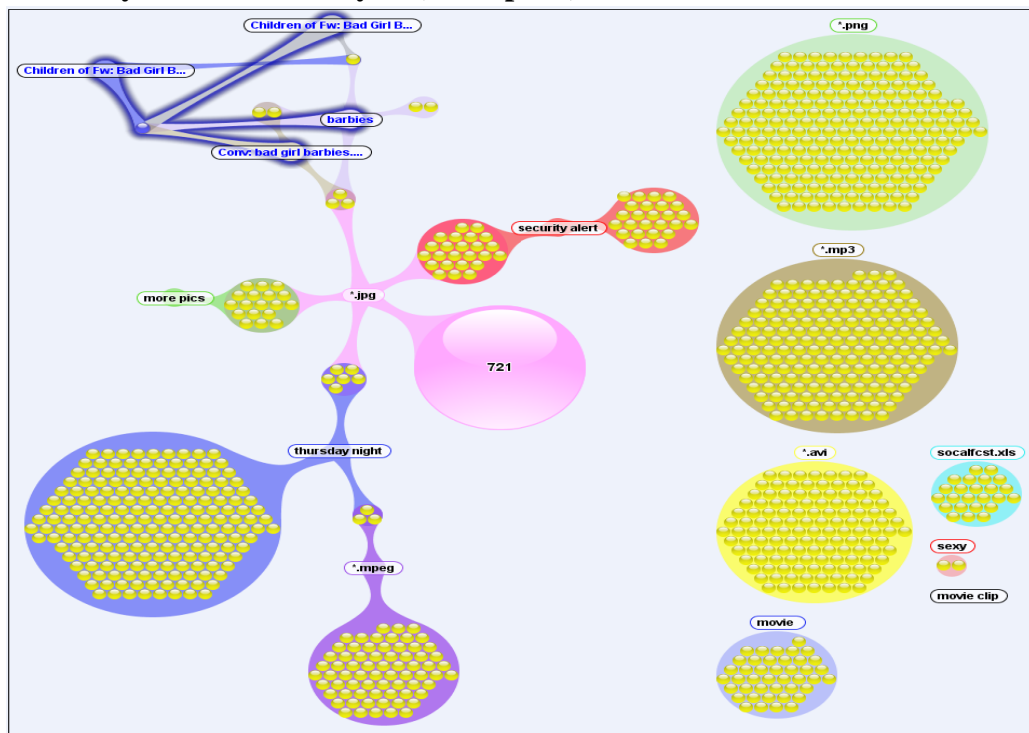
Appendix 6

e-Discovery Tool Analysis

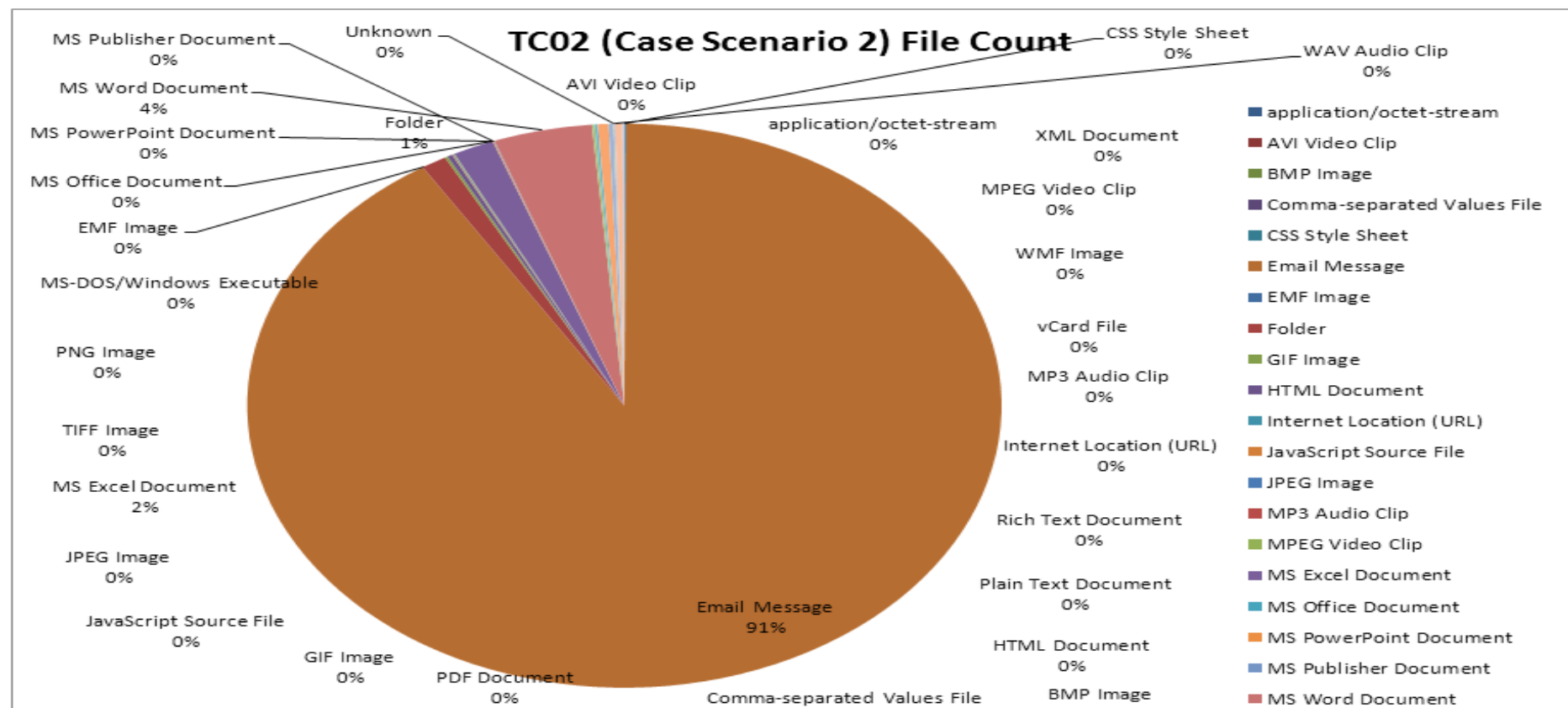
File / Keyword Link Analysis (Example 1)



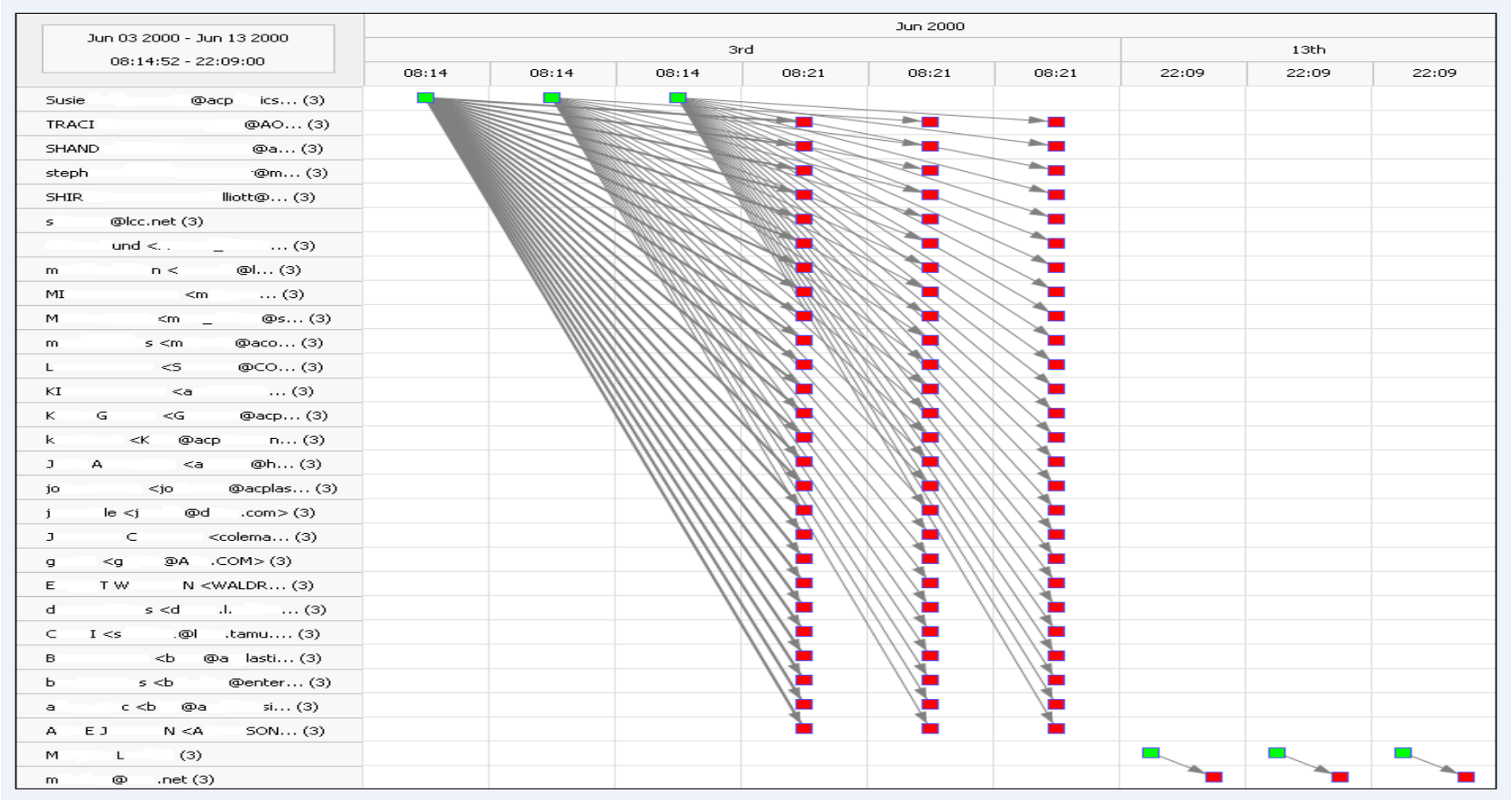
File / Keyword Link Analysis (Example 2)



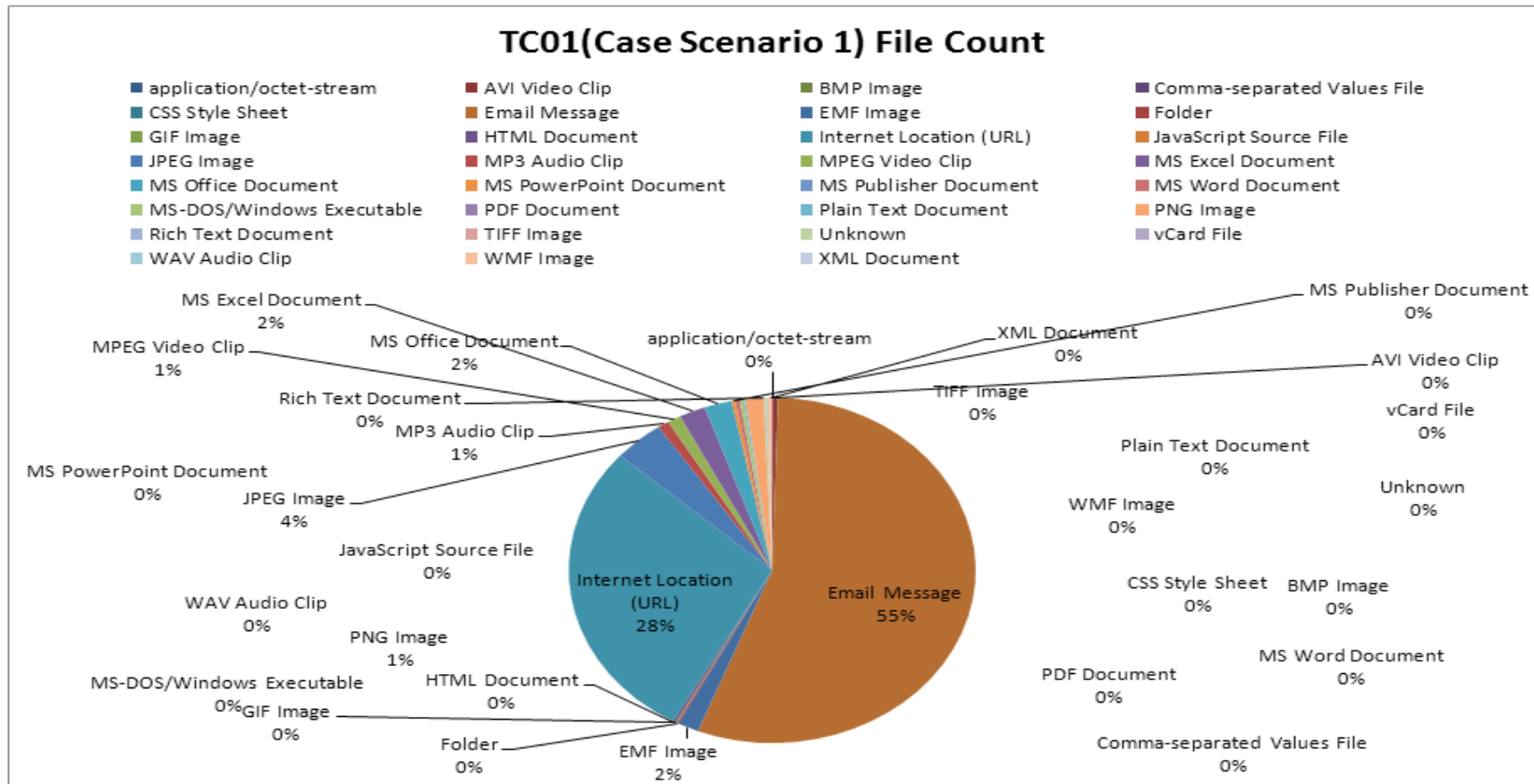
TC02 File Structure Analysis:



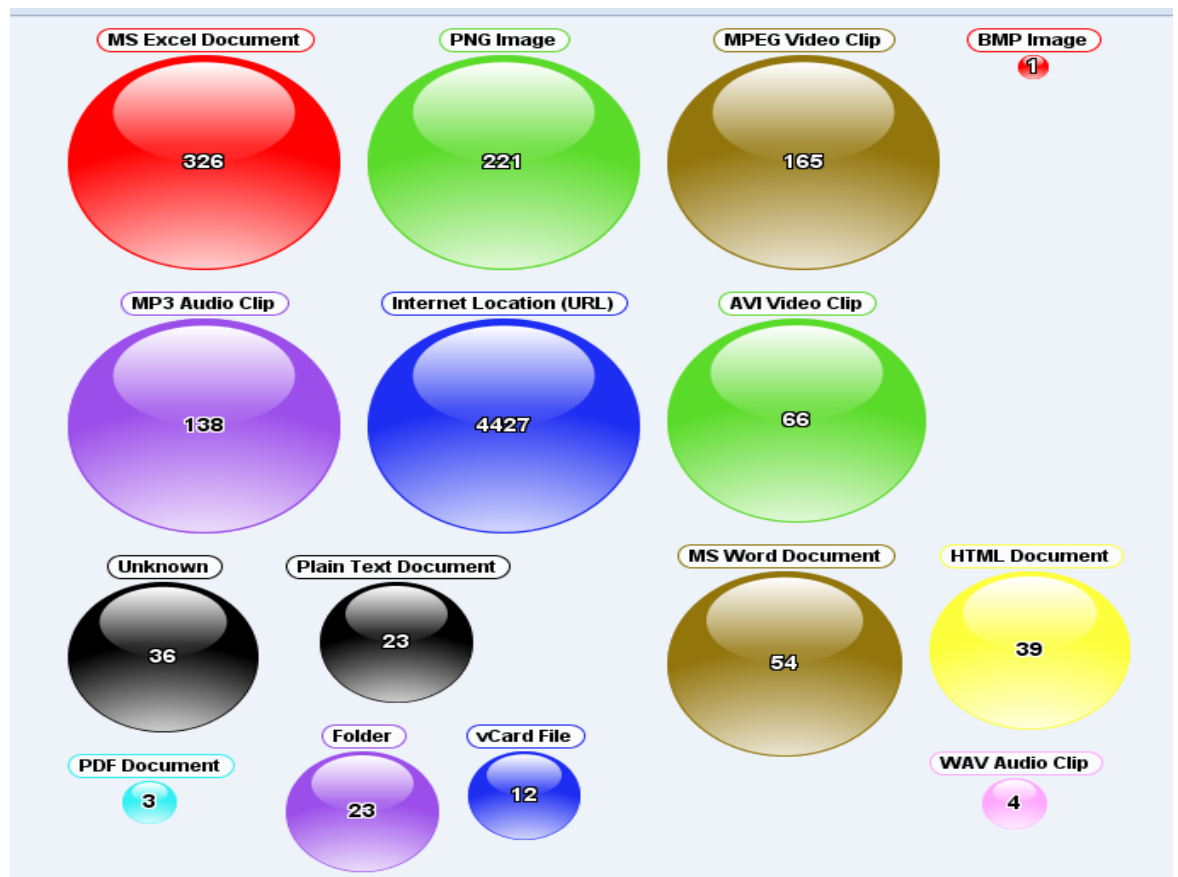
TC01 Email Conversation – Timeline:



TC-01 File Structure Analysis:



TC02-03 Search by Type



Appendix 7

e-Discovery Tool (Intella 1.5.2) Test Results

Case Scenario TC02 Email Conversation – Screen shot from Exported Report

Information Collected from Export Report File: (Case TC02) *XXX Original Information intentionally removed.

From: Paul <[REDACTED]@enron.com>
To: Susan <[REDACTED]@ENRON.com>
Subject: RE: [REDACTED]
Sent: 5 February 2002 4:35 AM
Received: 5 February 2002 4:35 AM

Subject: RE: [REDACTED]
Size: 2 KB (2,867 bytes)

Content Created: 19 June 2010 10:43 AM
Content Last Modified: 19 June 2010 10:43 AM

MIME Type: message/rfc822
Content MIME Type: text/plain
Character Set: utf8
Message Hash: e024d75b2e2fd94840 [REDACTED]
Message ID: <B9C853E5D [REDACTED] B148D@EULON-MSN>

Source: [REDACTED].pst
Location: [REDACTED].pst/Top of Personal Folders/[REDACTED]/Deleted Items

Headers

Date: Mon, 4 Feb 2002 07:35:52 -0800 (PST)
Message-ID: <B9C853E5DA596 [REDACTED] 1CB148D@EULON-MSMBX01V.corp.enron.com>
MIME-Version: 1.0
Content-Type: text/plain; charset=utf-8
Content-Transfer-Encoding: 7bit
Microsoft Mail Internet Headers Version 2.0: Microsoft Mail Internet Headers Version 2.0
Received: from nahou- [REDACTED].corp.enron.com ([192.168. [REDACTED] 0.2 [REDACTED]]) by NAHOU-MSMBX03V.corp.enron.com with Microsoft SMTPSVC(5.0.21 [REDACTED] 6); Mon, 4 Feb 2002 09:35:58 -0600
Received: from eulon- [REDACTED].corp.enron.com ([172.18. [REDACTED] 6 [REDACTED]]) by nahou- [REDACTED].corp.enron.com with Microsoft SMTPSVC(5.0.21 [REDACTED] 6); Mon, 4 Feb 2002 09:35:55 -0600
Received: from EULON-MSMBX01V.corp.enron.com ([172.18. [REDACTED] 89 [REDACTED]]) by eulon-mscnx04p.corp.enron.com with Microsoft SMTPSVC(5.0.21 [REDACTED] 6); Mon, 4 Feb 2002 15:35:52 +0000
content-class: urn:content-classes:message
X-MimeOLE: Produced By Microsoft Exchange V6.0.4712.0
Subject: RE: [REDACTED]
X-MS-TNEF-Correlator: <B9C853E5DA596 [REDACTED] 1CB148D@EULON-MSMBX01V.corp.enron.com>

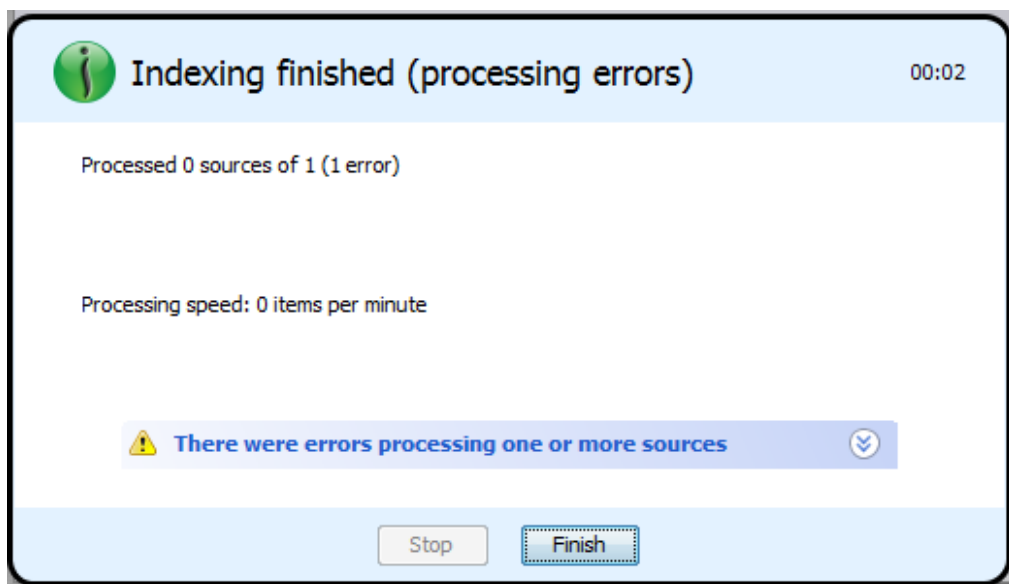
From: " Paul" <paul. xxxxxxxx @enron.com>
To: " Susan" <Susan. xxxxxxxx @ENRON.com>
Return-Path: paul. xxxxxxxx@enron.com
X-OriginalArrivalTime: 04 Feb 2002 15:35:52.0948 (UTC) FILETIME=[A xxxxxxxxxxxxxxxxxxxx1AD91]
X-ZL-From: xxxxxxxx , Paul
</O=ENRON/OU=NA/CN=RECIPIENTS/CN=EU/CN=RECIPIENTS/CN=P xxxxxxxxxx >
X-ZL-To: xxxxxxxx , Susan </O=ENRON/OU=NA/CN=RECIPIENTS/CN=S xxxxxxxxxx >
X-ZL-Subject: RE: Original Subject Removed
X-Filename: s xxxxxxxxx (Non-Privileged).pst
X-Folder: \Deleted Items
X-SDOC: 1331480
X-ZLID: zl-edrm-enron-v. xxxxxxxxxxxxxxxx 775.eml
X-ZL-Date: Mon, 4 Feb 2002 15:35:52 -0000

Findings / Recommendations [*Intella 1.5.2*]

Appendix 8

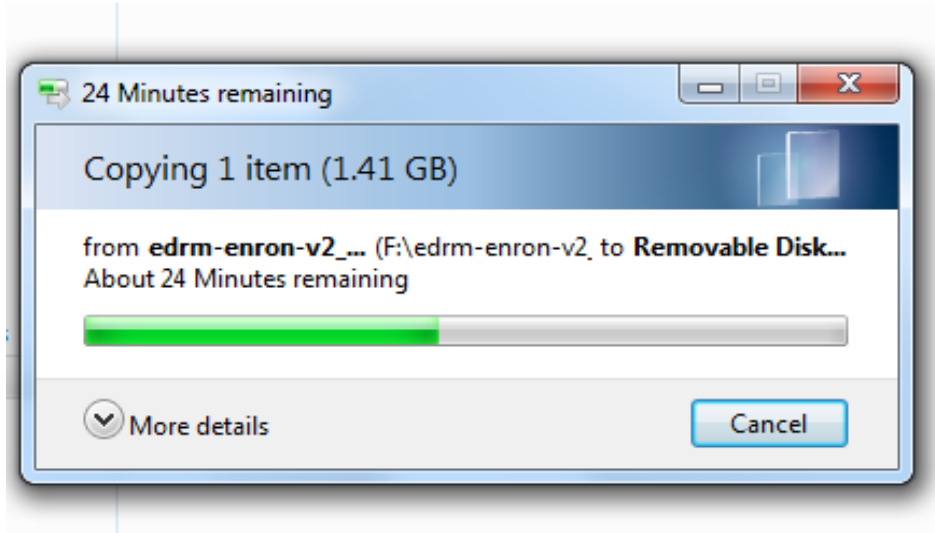
Test # 1 Error while processing zip file

Intella 1.5.2 failed to process / index a direct compressed (zipped) folder as a Source file. If you select the new source or new case with .Pst in zipped format it gives the following error.



Appendix 9


Test # 2 Unzipping PST File



Test Note: Therefore, it requires to Unzip the data and/or .Pst file – first and then select the Data set with .Pst format. This found as time consuming process. Above screenshot is the example of unzipping PST File.

Appendix 10

ESI Checklist: Ref: (www.edrm.net)

 ESI Checklist		Project/Case Ref: 01-02-03
<div style="text-align: right;">Print Form</div>		
File System Locations		
Common <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Local Hard Drives (operating system, size) <input checked="" type="checkbox"/> Laptop Hard Drives (operating system, size) <input type="checkbox"/> Network Share Hard Drives (operating system, size) <input checked="" type="checkbox"/> Mini-Device Storage Drives (operating system, size) <input checked="" type="checkbox"/> Deleted / Ambient Data (unallocated space) 	Less Common <ul style="list-style-type: none"> <input type="checkbox"/> Mainframes <input type="checkbox"/> Copy Machine Hard Drives <input type="checkbox"/> Scanning System Hard Drives <input type="checkbox"/> Printer Hard Drives <input type="checkbox"/> IM/Skype Archive system 	
User Removable Storage and Portable Devices		
<ul style="list-style-type: none"> <input checked="" type="checkbox"/> CD-ROMs <input checked="" type="checkbox"/> DVDs <input type="checkbox"/> Blu-Rays <input checked="" type="checkbox"/> USB Drives <input type="checkbox"/> MP3 Players/iPods <input checked="" type="checkbox"/> Removable Hard Drives 	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> External Hard Drives <input type="checkbox"/> PDAs <input type="checkbox"/> Mobile Phones/Devices <input type="checkbox"/> Digital Cameras <input type="checkbox"/> Removable Memory Cards <input type="checkbox"/> Digital Voice/Audio Recording Devices 	
Server Software		
<ul style="list-style-type: none"> <input type="checkbox"/> Network Operating System Information <input type="checkbox"/> Voice Mail <input type="checkbox"/> SharePoint Server <input checked="" type="checkbox"/> Local Email Stores <input type="checkbox"/> BlackBerry Enterprise Server (Mail, SMS, MMS, PIN) <input checked="" type="checkbox"/> Archiving Systems <input type="checkbox"/> Call Data Records (CDR) <input type="checkbox"/> Distribution List Membership Logs <input type="checkbox"/> Active Directory User Information <input type="checkbox"/> Video Surveillance Data/Logs <input type="checkbox"/> User Activity/Monitoring Logs (e.g. Web Trends, Golden Eye, etc.) <input type="checkbox"/> Mail Server (Servers/routers installed & whether/how users remotely access) 	<ul style="list-style-type: none"> <input type="checkbox"/> Proprietary Systems <input type="checkbox"/> Inventory Systems <input type="checkbox"/> Procurement Systems <input type="checkbox"/> HR Systems <input checked="" type="checkbox"/> Database Applications <input type="checkbox"/> Finance/Audit Systems <input checked="" type="checkbox"/> Document Management Systems <input type="checkbox"/> CRM Systems <input type="checkbox"/> Accounting Systems <input type="checkbox"/> Remote Computer Connection Systems 	
Public/Semi-Shared Sources		
<ul style="list-style-type: none"> <input type="checkbox"/> Social Networking <input type="checkbox"/> Blogs <input type="checkbox"/> Web Pages <input type="checkbox"/> Wikis <input type="checkbox"/> Any other software where company has possession, custody or control 	<ul style="list-style-type: none"> <input type="checkbox"/> Public Domain Forums <input type="checkbox"/> Web Storage (e.g. Drop Box) <input type="checkbox"/> Cloud Computing Systems <input type="checkbox"/> Third Party DM Systems (e.g. Google Docs) 	
Security/Access Components		
<ul style="list-style-type: none"> <input type="checkbox"/> Passwords <input checked="" type="checkbox"/> Encryption/Decryption Keys (e.g. .ID files for Notes) <input type="checkbox"/> Website access/usage records 		
Backups		
<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Active System <input type="checkbox"/> Disaster Recovery <input type="checkbox"/> Tapes (retired or unidentified) 		
Retired Workstations, Mail Servers, Devices		
<ul style="list-style-type: none"> <input type="checkbox"/> Legacy System Data 		

KEY: ☒ Include ☐ Exclude ☐ Not Determined

E-mail comments & suggestions to: ESIChecklist@EDRM.net OR [Click here to post a comment on our web site](#)

EDRM ESI Checklist Ver. 1.00e jd 01/07/2011 Page 1 of 1

Appendix 11

The e-Discovery tool (*Intella 1.5.2*) Features / Functionality (www.vound-software.com).

[Test - Checklist]

Intella 100 GB - Features	Product Feature	Product Testing
Case Options		
Create Case	✓	☑
Case Export	✓	☑
Max Case Size	100 GB	☑
Multi User Cases	N/A	N/A
File Search Support		
MS Office	✓	☑
PDF	✓	☑
Images	✓	☑
Most file types	✓	☑
Search Tools		
Search Facets	✓	☑
Tagging	✓	☑
Item Preview	✓	☑
Cluster Map	✓	☑
Table View	✓	☑
Thumbnail View	✓	☑
Timeline View	✓	☑
Tree View	✓	☑
Word Lists	✓	☑
De-duplication		
Files	✓	☑
E-mail	✓	☑
Search Options		
Keyword	✓	☑
Proximity	✓	☑
Wildcards	✓	☑
Fuzzy Search	✓	☑
MD5 / Message Hash	✓	☑
Keyword List	✓	☑
MD5 List	✓	☑
Export Options		
Native export	✓	☑
PDF	✓	☑

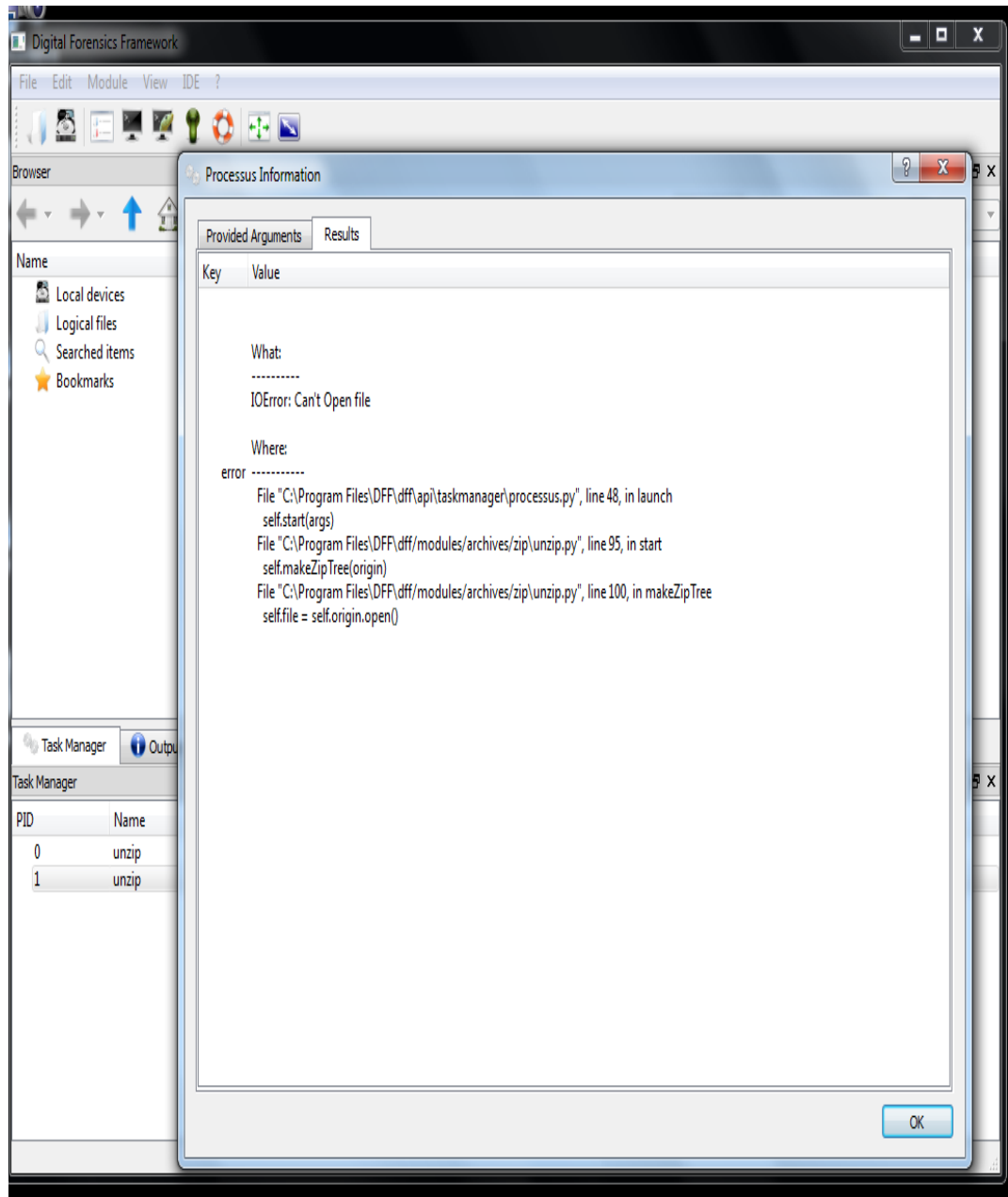
Outlook PST	✓	☑
File Numbering	✗	✗
PDF Footer Numbering	✗	✗
Word Lists	✓	☑
E-mail Support		
Lotus Notes	✓	✎
Novel GroupWise	✓	✎
Outlook PST	✓	☑
Outlook - OST	✓	☑
Outlook Express -EML	✓	✎
WinMail	✓	✎
Foxmail	✓	✎
Thunderbird - Mbox	✓	✎
Eudora	✓	✎
Reporting		
HTML Report	✓	☑
CSV	✓	☑
PDF, RTF	✓	☑
Note / Remark		
Product Features	✓	
Not Supporting	✗	
To be tested	✎	
Tested - OK	☑	

Appendix B

(Open-Source Tool – Digital Forensics Framework (DFF))

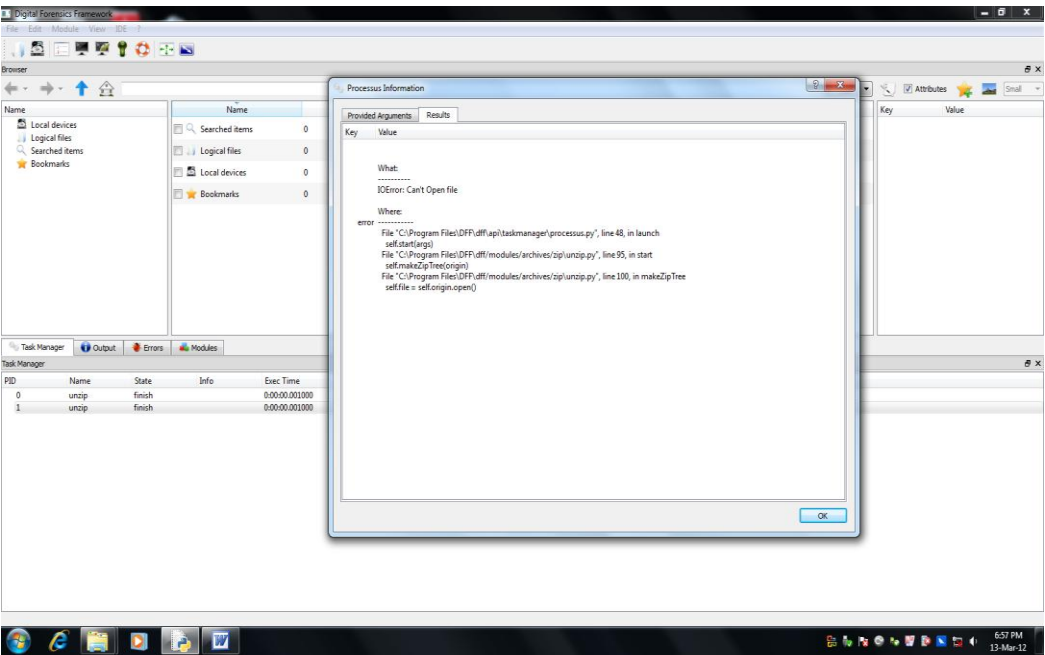
DFF – Pilot Test Processing & Findings.

Test # 1 PST file Processing - Unsuccessful

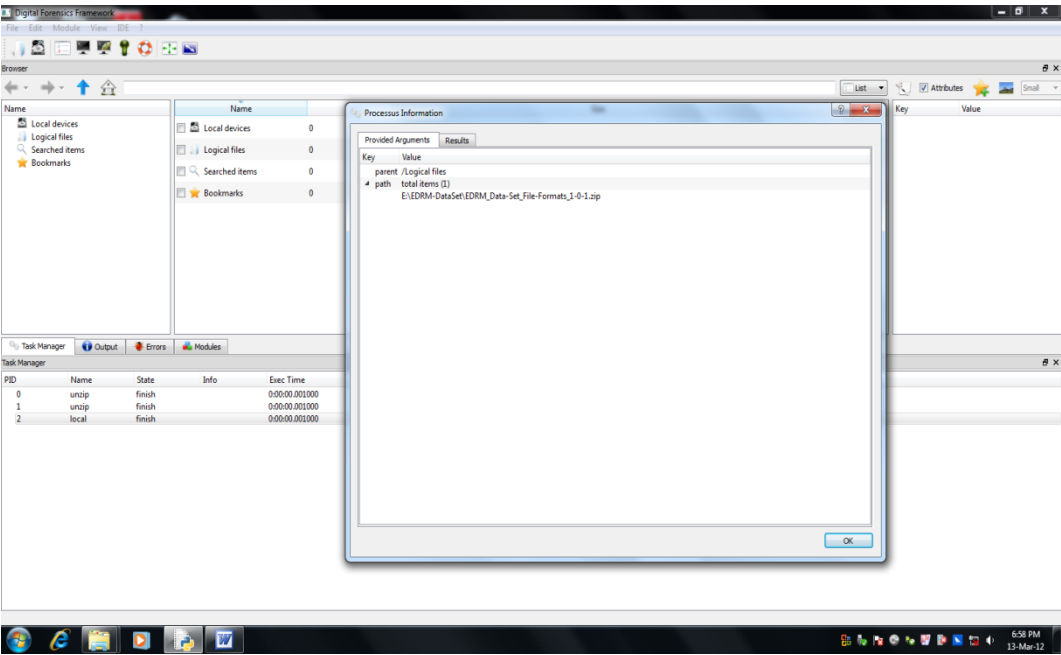


Test Note 1: The Arnold (zl_arnold-j_000) .pst file was not successfully processed by the open-source tool on the first attempt.

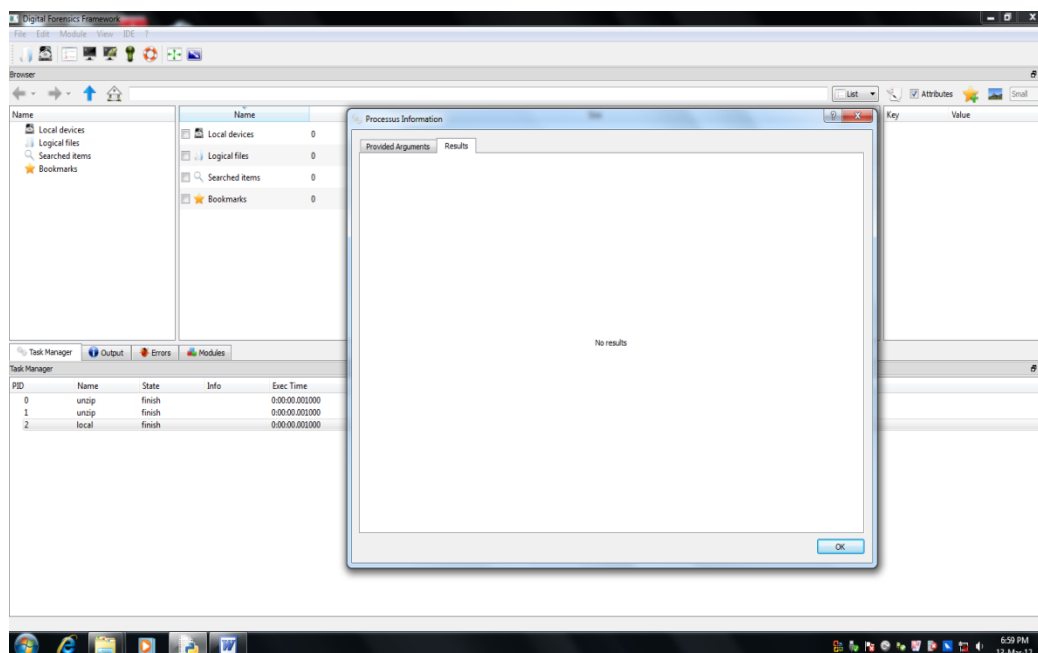
Test # 2 EDRM Dataset File Format Processing - Unsuccessful



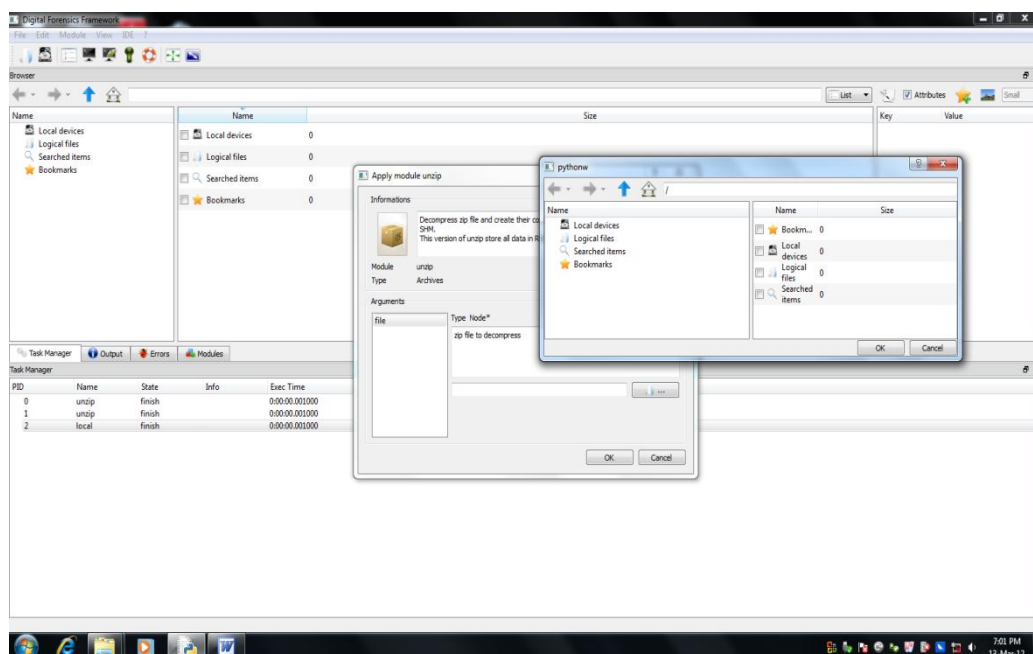
Test Note 2: Second test failed with EDRM Dataset File format.



Test # 3 Opening Zip files - Successful

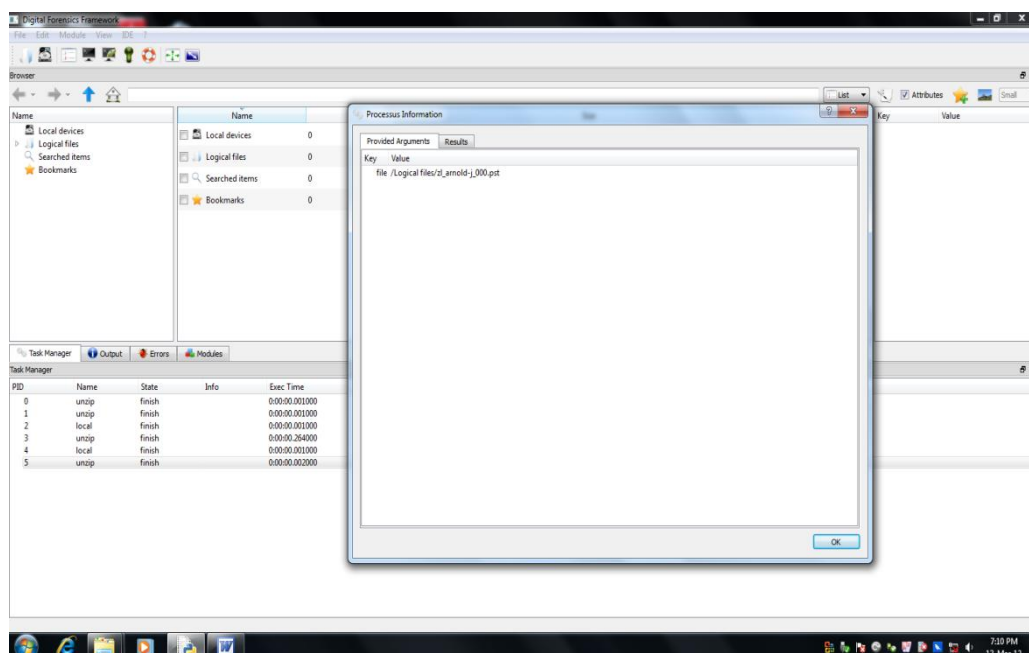
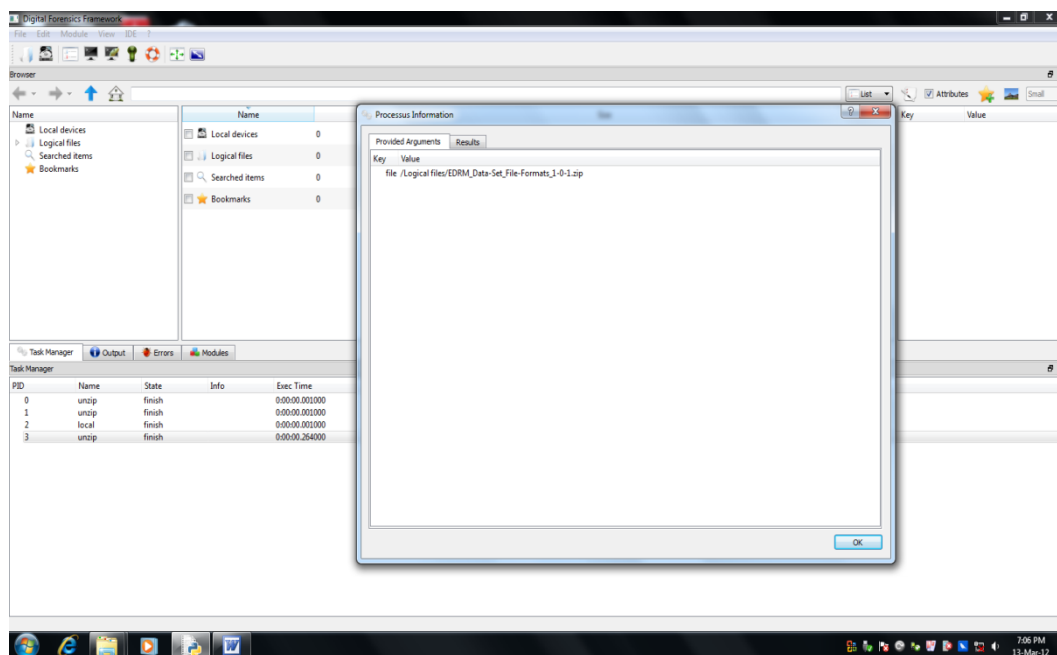


Test Note 3: To open zip files for non-technical users was complex.



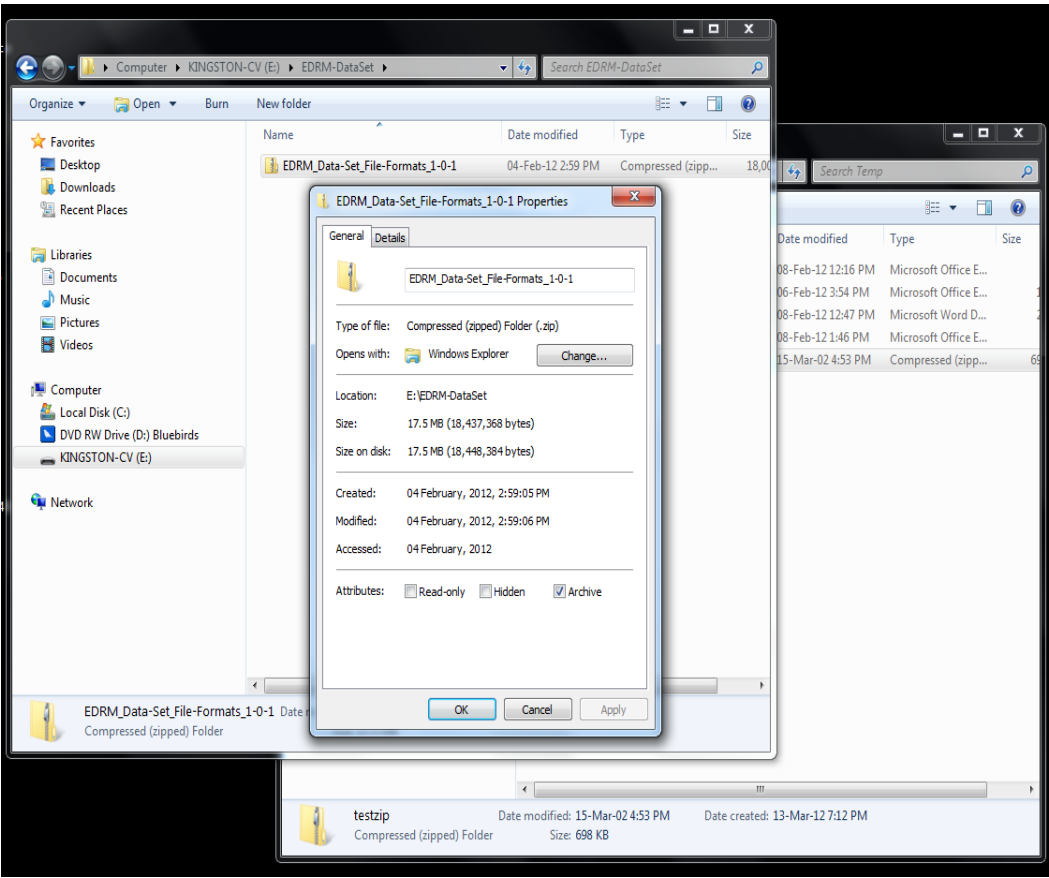
Test Note 4: Successful – Unzip process time 0:00:00.264000.

Test # 4 PST File Processing - Success



Test Note 5: The Arnold (zl_arnold-j_000) .pst file successfully processed through open- source tool on the fourth attempt.

Test # 5 EDRM – Data size Details from Windows Property



Test # 6 File Search – Complexity

The screenshot shows the Digital Forensics Framework (DFK) interface. The main window displays the 'hevedit_dj_000.pst' file. The search results pane on the right shows a search for 'Character(s)' with a pattern of 'jpg' and a wildcard of '*.jpg'. The search results table at the bottom lists various tasks and their execution times.

PID	Name	State	Info	Exec Time
0	unzip	finish		0:00:00.001000
1	unzip	finish		0:00:00.001000
2	local	finish		0:00:00.001000
3	unzip	finish		0:00:00.264000
4	local	finish		0:00:00.001000
5	unzip	finish		0:00:00.002000
6	local	finish		0:00:00.001000
7	carver-gui <Logical files>	wait		0:29:39.004000
8	hash	finish		0:00:00
9	viewimage	wait		0:24:25.067000
10	metaeif	finish	Registering node /	0:00:00
11	hevedit_dj_000.pst	wait		0:07:34.013000

Task Manager				
Task Manager				
PID	Name	State	Info	Exec Time
0	unzip	finish		0:00:00.001000
1	unzip	finish		0:00:00.001000
2	local	finish		0:00:00.001000
3	unzip	finish		0:00:00.264000
4	local	finish		0:00:00.001000
5	unzip	finish		0:00:00.002000
6	local	finish		0:00:00.001000
7	carver-gui <Logical files>	wait		0:29:39.004000
8	hash	finish		0:00:00
9	viewerimage	wait		0:24:25.067000
10	metaexif	finish	Registering node: /	0:00:00
11	hexedit zl_arnold-j_000.pst	wait		0:07:34.013000

Test Note 6: This screenshot shows the “Task Manager” list with execution timing, status of processing and file name from the previous test (Test #6) File Search – a complex and time-consuming process.