Hochschule Wismar
University of Applied Science,
Technology, Business and Design
Fakultät für Ingenieurwissenschaften
Bereich Elektrotechnik und Informatik

Master-Thesis

Untersuchungen zum Einsatz von eDiscovery-Systemen für IT-forensische Zwecke in Strafverfahren

Abschlussarbeit zur Erlangung des akademischen Grades

Master of Engineering (M.Eng.)

im Studiengang IT-Sicherheit und Forensik

Erstgutachter:	Frau Prof. Dr. Ing. Antje Raab-Düsterhöft
Weitere Gutacher:	Herr Dipl.-Ing. Hans-Peter Merkel
	Herr Dipl.-Ing. Gilbert Löhr
eingereicht von:	Frank Meixelsperger geboren am 28.08.1964 in Bonn
Matrikelnummer:	328070
Datum der Abgabe:	16.02.2023

Vorwort und Danksagung

Trotz meiner langjährigen Praxiserfahrung und Sachverständigen - Ausbildung beim BKA hat sich immer wieder gezeigt, dass der Erfolg zur Lösung von neuen Aufgabenstellungen in dieser Disziplin eindeutig vom Grundlagenwissen in der Informatik abhängt. Gerade dann, wenn es in diesem Themenfeld Fragen zu beantworten gilt, die eben nicht mit standardisierten Untersuchungsmethoden oder Software-Trainings von IT Forensik Tool-Herstellern lösbar sind, ist es entscheidend, ob man eigenständig neue Lösungswege zur Beantwortung aufgrund seines Basiswissens finden und umsetzen kann oder nicht. Das gelingt meiner Beobachtung nach Ingenieuren mit einem wissenschaftlichen Studienabschluss im Vergleich zu einem „angelernten“ Kriminalbeamten wie mir besonders gut. Grund genug für mich, sich diesem Studium und seinen Anforderungen zu stellen.

Bedanken möchte ich mich bei Frau Prof. Dr. Ing. Antje Raab-Düsterhöft und Herrn Diplom Ingenieur Hans-Peter Merkel, die mir die Chance gegeben haben, dieses Studium überhaupt aufnehmen zu dürfen.

Eine der grundlegenden Aufgaben des Bundesamtes für Sicherheit in der Informationstechnik ist es Leitlinien und Handlungsempfehlungen in der IT-Sicherheit zu fertigen. Der Leitfaden „IT-Forensik“ des BSI aus dem Jahre 2011, den ich im Rahmen meines Studiums gelesen habe, hat nach wie vor trotz seines Alters inhaltliche Bestandskraft und zeigt mir die Qualität der Arbeit. Dies hat mich veranlasst als weiteren Betreuer Herrn Gilbert Löhr vom BSI zu gewinnen, da ich bei meinen Recherchen zum Themengebiet eDiscovery keine deutschsprachige Grundlagenarbeit hierzu finden konnte und der erste Teil meiner Arbeit aus der grundlegenden Funktionsbeschreibung eines eDiscovery-Systems besteht.

Vielen Dank an meine Betreuer für die gewährte Unterstützung!

Frank Meixelsperger

Aufgabenstellung

Electronic-Discovery Systeme werden nicht nur in ihrem ursprünglichen Anwendungsgebiet bei Wirtschaftsprüfungsgesellschaften und internationalen Anwaltskanzleien, sondern zunehmend auch in Sicherheitsbehörden eingesetzt.
Dabei ist zu beobachten, dass nicht alle Länderpolizeien diesem Trend folgen. Gründe hierfür sind die Komplexität, Fehlerhaftigkeit einiger Systeme und hohen Anschaffungskosten. Ebenfalls erkennbar ist, dass die ablehnenden Länderpolizeien keine eigene differenzierte Meinung entwickeln, häufig aus Mangel an Zeit für aufwendige Tests und dem lang anhaftenden Ruf der Funktionsuntüchtigkeit dieser Systeme für IT -forensische Zwecke.
Mittlerweile hat aus Sicht des Autors das führende eDiscovery-System eine Funktionsreife erlangt, die die hohen Anschaffungskosten rechtfertigt. Belegt wird diese Annahme durch die Tatsache, dass Sicherheits- und Kontrollbehörden des Bundes und der Länder seit einigen Jahren umfänglich mit dieser Anwendung arbeiten, trotz hoher Finanzmittel- und Personalaufwendungen.

Diese Master-Thesis soll dazu beitragen, das Prinzip von eDiscovery-Systemen verständlicher zu machen und so dem Leser eine fundierte Entscheidungsgrundlage für eine mögliche Einführung in seiner Sicherheitsbehörde zu liefern.

Dazu werden im ersten Teil die Grundlagen und die prinzipielle Funktionsweise von eDiscovery-Systemen erklärt, um auch die Gründe für die Fehlerhaftigkeit bzw. Mängel verständlich zu machen.

Im zweiten Teil der Master-Thesis wird untersucht wie weit diese Verarbeitungsgebnisse von eDiscovery-Anwendungen die Ansprüche an Beweismittel in Strafverfahren bezüglich Qualität und Authentizität erfüllen oder Risiken beinhalten und ob sie geeignet sind, bestehende klassische IT-Forensik Arbeiten vollständig abzulösen.
Hierfür werden mehreren Testszenarien mit zwei bekannten eDiscovery-Tools und der Untersuchung dieser Ergebnisse durchgeführt und am Ende eine abschließende Bewertung und Empfehlung abgegeben.

Kurzreferat

Electronic Discovery-Systeme werden wegen ihres hohen Automationsgrades in der Aufbereitung digitaler Daten, deren zentraler Datenhaltung und Netzwerkfähigkeit sowie der zeitsparenden und damit effektiven Auswertungsmethode durch Ermittler vor allem bei Großverfahren mit vielen Tatbeteiligten oder großen Datenmengen in deutschen Sicherheitsbehörden immer häufiger eingesetzt und lösen zunehmend klassische Arbeitsmethoden und -abläufe in der IT-forensischen Untersuchung und Datenaufbereitung ab.
Zudem entsteht ein Mehrwert durch den Vergleich gleichartiger Datentypen aus unterschiedlichen Datensicherungen verschiedener Täter und der Visualisierung dieser Big Data-Analyseergebnisse. So können z.B. über die Analyse von E-Maildaten oder Telefonnummern aus unterschiedlichen Datensicherungen die Kommunikationsbeziehungen und auch ihre Häufigkeit zwischen Tätern dargestellt werden, was für das Erkennen einer Bandenstruktur und der Rollenverteilung innerhalb dieser Gruppe wichtige Erkenntnisse für die Ermittlungen bedeutet.

Im Gegensatz zur klassischen Arbeitsweise in der IT-Forensik wird hier ein stark automatisierter Verarbeitungsprozess zur Aufbereitung digitaler Beweismittel eingesetzt. Dabei wird die Datenaufbereitung technisch so verändert, dass es zur Reduktion der zu untersuchenden Datenmenge und zu einer einfacheren, schnelleren und damit effektiveren Methode des Suchens bzw. Recherchierens kommt.

Ziel dieser Masterarbeit ist die Nutzbarkeit von eDiscovery-Systemen als neues Instrument in der IT-forensischen Untersuchungsarbeit zu prüfen und zu bewerten.

Abstract

Due to their high degree of automation in the processing of digital data, their central data storage and network capability, as well as the time-saving and therefore effective evaluation method by investigators, electronic discovery-systems are being used more and more frequently in German security authorities, especially in large-scale proceedings involving many criminals or large amounts of data, and are increasingly solving classic working methods and processes in IT forensic investigation and data processing.

In addition, added value is created by comparing similar data types from different data backups from different perpetrators and the visualization of these Big Data analysis results. For example, by analyzing e-mail data or telephone numbers from different data backups, the communication relationships and their frequency between suspects can be shown, which means important insights for the investigations in order to identify a gang structure and the distribution of roles within this group.

In contrast to the classic way of working in IT forensics, a highly automated processing process is used here to prepare digital evidence. The data processing is technically changed in such a way that the amount of data to be examined is reduced and a simpler, faster and thus more effective method of searching or researching is achieved.

The aim of this master's thesis is to test and evaluate the usability of eDiscovery-systems as a new instrument in IT forensic investigation work.

Inhaltverzeichnis

1 Einleitung

1.1 Motivation

Die Idee zur Entwicklung von electronic Discovery-Systemen entstand aus dem Bedarf, eine kostengünstigere Methode zu entwickeln, um die aufwendige und teure manuelle Sichtung, Überprüfung und Bereitstellung von Unternehmensdokumentationen durch Anwälte für US Zivilklageverfahren durch Automation zu minimieren und so die Unkosten zu senken. Das US-Zivilrecht schreibt den Austausch von Unterlagen mit möglicher Verfahrensrelevanz zwischen den beteiligten Prozessparteien vor. Die Verweigerung der Herausgabe oder Unterschlagung von verfahrensrelevanten Informationen kann erhebliche finanzielle Nachteile für den Beschuldigten nach sich ziehen. Derartige Dienstleistungen von Anwaltskanzleien machen häufig den höchsten Unkostenanteil im Prozess aus.

Durch die zunehmende Digitalisierung standen Unterlagen und Kommunikation wie E-Mails oder Buchhaltung der Firmen nicht mehr nur in papierner, sondern auch digitaler Form zur Verfügung. Damit entstand die Idee, diese Daten unterschiedlichster Art über Programmierung zusammenzuführen, diese vielfältigen Datenformate zu harmonisieren und mittels einer indexierenden Datenverarbeitungsform recherchefähig zu machen, was zu großer Zeit- und damit Kostenersparnis führte. Denn die Gesamtheit aller Dokumente musste nicht mehr einzeln gesichtet werden, um relevante Inhalte zu finden, sondern nur noch die jeweiligen Suchtreffer aus den Abfragen. Die über Suchtreffer festgestellten verfahrensrelevanten Informationen werden dann an den Prozessgegner übergeben.

Mittlerweile haben sich eDiscovery-Systeme bei vielen internationalen Anwaltskanzleien, Compliance-Abteilungen internationaler Aktiengesellschaften, dem Surpreme Court der USA, dem Internationalen Strafgerichtshof in Den Haag aber auch bei internationalen Wirtschaftsprüfungsgesellschaften wie KPMG, PWC oder Ernst & Young etabliert.

Auch Zeitungsverlage entdecken für sich die besonderen Fähigkeiten von eDiscovery-Systemen. So wurden durch Hacking erlangte Kundendaten aus einer auf Offshore-Kapitalanlagen spezialisierten panamaischen Anwaltskanzlei der Washington Post

und der Süddeutschen Zeitung zu Verfügung gestellt. Wegen der Menge und Unübersichtlichkeit der Informationen wurden diese Daten in einem eDiscovery-System in Deutschland recherchefähig und über gesicherte Datenleitungen weltweit Investigativ-Reportern verfügbar gemacht. Die Veröffentlichungen hierzu sind besser bekannt als „Panama und Paradise Papers“.

Dieses Grundprinzip, alle elektronischen Daten zusammenzuführen, zu indexieren und unter einer einheitlichen Bedienoberfläche auszuwerten, lässt sich ebenso auf die Unterstützung bei polizeilicher Auswertung von digitalen Beweismitteln übertragen.

1.2 Ausgangssituation und Problemstellung

Kriminalfälle wurden in den Anfängen des letzten Jahrhunderts häufig durch Vernehmungen von Tatverdächtigen und Zeugen, der Überprüfung deren Aussagen und logischen Schlussfolgerungen versucht aufzuklären. Das heißt, staatsanwaltschaftliche und richterliche Entscheidungen stützten sich damals vorwiegend auf Personalbeweise. Personalbeweise sind häufig subjektiv, denn Tatverdächtige wie auch Zeugen können aus verschiedenen Gründen die Unwahrheit sagen und nicht immer lässt sich der Wahrheitsgehalt eindeutig überprüfen.

Durch die Entstehung und Weiterentwicklung der Kriminaltechnik werden heute Urteile auf der Basis von Sachbeweisen wie z.B. gerichtsmedizinischen, daktyloskopischen oder DNA-Spuren, soweit vorhanden, gefällt. Sachbeweise sind im Gegensatz zum Personalbeweis objektiv, neutral, unbeeinflussbar und bilden damit eine sichere Urteilsgrundlage für das Gericht.

Digitale Beweismittel zählen zu diesen Sachbeweisen und nehmen mittlerweile einen großen Raum in der Beweisführung ein. Daraus entstand in den letzten 20 Jahren eine neue kriminaltechnische Disziplin - die IT-Forensik. Da die heutige Gesellschaft in jeder Form digitale Daten für ihr Leben nutzt, lassen sich daraus viele Erkenntnisse bzw. Spuren für polizeiliche Ermittlungen und spätere Gerichtsverhandlungen gewinnen und sind mittlerweile unverzichtbar in der polizeilichen Ermittlungsarbeit.

Die Tatsache, dass digitale Beweismittel häufig zur Fallaufklärung beitragen oder einen ersten Ermittlungsansatz liefern können, führte dazu, dass der Bedarf bei den

Ermittlern auf IT-forensische Untersuchungsarbeiten immer weiter anstieg. Zugleich stieg auch die Datenmenge und Vielfältigkeit der Spurenquellen je Fall an und damit der technische Aufwand, um an diese Daten zu gelangen und sie für die Ermittler in auswertbare Form zu bringen.

Dadurch bedingt müssen die Ermittler immer mehr Zeitanteile für die Auswertung aufwenden und unterschiedlichste Auswertetools handlungssicher bedienen können, um die verfahrensrelevanten Spuren in den aufbereiteten Datenextrakten der IT-Forensiker zu finden. Zudem erschwert die schiere Menge an digitalen Daten die Übersicht zu behalten und auch Bezüge zwischen den einzelnen Tätern zu erkennen, beispielsweise durch den Abgleich hunderter Telefonbucheinträge aus Smartphones oder Instant Messenger-Kommunikation verschiedener Tatbeteiligter.

1.3 Zielsetzung und Aufgabenstellung

Electronic Discovery-Systeme können dazu beitragen, die oben dargestellten Probleme zu mindern, indem sie

- dem Auswerter eine einzige einheitliche Bedienoberfläche zur Verfügung stellen, denn die Konzentration auf nur noch eine statt mehrere Bedienoberflächen erhöht die Handlungssicherheit und damit die Chance, effektiv die verfahrensrelevanten Informationen zu finden

- den Auswerter bzw. Ermittler von dem Management der unterschiedlichen Datenträger (USB-Sticks, -Festplatten, Blu-ray´s oder DVD´s) befreien und die Gesamtheit aller in einem Vorgang erstellten digitalen Datenextrakte zentral auf leistungsstarken Fileservern in einem Netzwerk zur Verfügung stellen

- die Datenmenge durch verschiedene Techniken und Funktionen von eDiscovery-Systemen reduzieren, um den Auswerteaufwand zu minimieren und damit effektiver zu gestalten

- die Auswertung über ein zentral gesteuertes Client-Server-Netzwerk durch parallelen Zugriff und damit schnellere Auswertung zu unterstützen

- Gleichheiten bei Kommunikationsdaten (Rufnummern, Nicknames, E-Mail-Adressen, ISM-IDs etc.) in Datensicherungen von unterschiedlichen Tätern im gemeinsamen Verfahren finden und so mögliche Beziehungen visualisieren.

Die ursprüngliche Zielsetzung bei der Entwicklung von eDiscovery-Systemen konzentrierte sich vor allem auf bekannte Datenformate und deren Aufbereitung weit verbreiteter Softwareanwendungen in Unternehmen. Dies umfasst vor allem digitalisierte Informationen zur Kommunikation wie E-Mailprogramme, Office-Anwendungen wie Tabellenkalkulation, Textverarbeitung und Buchhaltung, die für einen Zivilprozess Relevanz besitzen.

EDiscovery-Systeme für Strafverfolgungszwecke müssen dagegen nicht nur diese Ursprungsziele erfüllen, sondern auch Sicherungsformate aus der IT-Forensik-Welt, Partitionen und unterschiedliche Dateiverwaltungssysteme erkennen und extrahieren sowie Daten aus faktisch allen möglichen Anwendungen, die ein Tatverdächtiger nutzen kann, sichtbar machen können.

Diese Masterarbeit untersucht anhand verschiedener Funktionen innerhalb der Verarbeitungs- und Erzeugungsprozesse von zwei aktuellen eDiscovery-Programmen, inwieweit das eDiscovery-Prinzip für IT-forensische Aufgaben in deutschen Strafverfahren geeignet ist und welche Risiken sich im Hinblick auf Gerichtsfestigkeit bzw. -verwertbarkeit dieser Verarbeitungsergebnisse ergeben können.

1.4 Abgrenzung

In verschiedenen Electronic Discovery Reference Modell - Handlungsempfehlungen bzw. Richtlinien (z.B. Produktionsleitfaden [1]) geht das EDRM - Gremium immer wieder auf die auszuhandelnden Absprachen zwischen den Prozessparteien in US – Zivilverfahren ein. Hierbei geht es um vereinbarte Exportformate oder festgelegte Arbeitsschritte für den schrittweisen Datenaustausch. Dieses Thema und der letzte Prozessschritt im eDiscovery-Prozessmodell, genannt Präsentation, werden für die Frage nach einer Geeignetheit in deutschen Strafverfahren nicht benötigt und daher nur kurz, aber nicht näher ausgeführt.

[1] https://edrm.net/resources/frameworks-and-standards/edrm-model/production/

Diese Masterarbeit befasst sich im Kern mit dem technischen Verarbeitungsprozess und seinen Ausgabeformaten.

1.5 Vorgehensweise

1.5.1 Funktionsweise eines eDiscovery-Systems

Bei den Recherchen zu dieser Master-Thesis wurde kein deutsch- oder englischsprachiges Standardwerk für eDiscovery-Systeme gefunden. Aufgrund der Komplexität derartiger Systeme werden in diesem Abschnitt die Grundlagen und Funktionalitäten eines eDiscovery-Systems tiefergehend vorgestellt, um so das Verständnis für Problemstellungen und den dazu durchgeführten Untersuchungen und deren Ergebnisse verständlicher zu machen.

1.5.2 Untersuchungen zu eDiscovery-Systemen

Die Untersuchungen bestehen aus zwei Fragestellungen:

1. Erfüllen eDiscovery-Systeme die Ansprüche an IT-Forensische Arbeiten?
   Hierzu wird ein formaler und technischer Vergleich der Merkmale zwischen dem erklärten Electronic Reference Discovery-Modell (EDRM) [2] und dem bekannten SAP-Modell gezogen.

2. Wird das originale digitale Beweismaterial durch den Verarbeitungsprozess eines eDiscovery-Systems so verändert, dass seine Nutzbarkeit für deutsche Strafverfahren Risiken birgt?
   Anhand von Testszenarien mit zwei eDiscovery-Anwendungen wird dies überprüft.

1.5.3 Zusammenfassung und Bewertung

Die Untersuchungsergebnisse werden mit den Anforderungen an klassische IT-Forensik Prozesse verglichen und ein Resümee im Hinblick auf seine Tauglichkeit und Einsatzgebiete für Strafverfahren gezogen.

[2] www.edrm.net

2 Grundlagen eines eDiscovery-Systems

Es gibt keine einheitlichen Standards für eDiscovery-Anwendungen. Workflows und Best Practices können von Anbieter zu Anbieter unterschiedlich sein.
Als Orientierung und weltweit anerkannter Leitfaden gilt allerdings das Electronic Discovery Reference Model (EDRM), welches im nachfolgenden näher erläutert wird und als Ausgangsbasis für die späteren Untersuchungen dient.

2.1 Definition eines eDiscovery-Modells

2005 gründeten George Socha und Tom Gelbmann die EDRM Community, eine Gruppe von Rechts- und eDiscovery-Experten, die 2016 von der Duke Law School in den USA und im Oktober 2019 von Mary Mack und Kaylee Walstad übernommen wurden. Die dort erstellten Definitionen, Leitlinien und Anforderungen an ein eDiscovery-System gelten heute als weltweiter Branchenstand und werden in diesem Gremium immer weiterentwickelt.

Die Gremiumsarbeit des EDRM soll - ähnlich wie bei dem RFC (Request for Comments = Gremium für Internetstandards) - zur Standardisierung von eDiscovery-Systemen führen, um so dazu beizutragen, gleiche Bedingungen zwischen unterschiedlichen Anbietern aber auch Prozess-Parteien beim Austausch von Daten zu schaffen und die Prozesse zu vereinfachen. Inhalte bzw. Bezüge zur IT-forensischen Anwendbarkeit in Strafverfahren wurden in den dortigen Dokumenten zum Zeitpunkt der Erstellung dieser Master-Thesis nicht gefunden.

Im EDRM-Dokument „Introduction“ [3] wird ein eDiscovery-System folgendermaßen definiert:

- *E-Discovery beinhaltet den Austausch, die Analyse und die Überprüfung von elektronischen Dateien, E-Mails und anderen Informationen, die auf einem Computergerät gespeichert sind.*

- *Das Ziel einer eDiscovery-Anwendung ist es, relevante Informationen für ein Gerichtsverfahren, ein Schiedsverfahren oder eine Anhörung aufzudecken. Der*

[3] https://edrm.net/wiki/introduction/

Prozess beginnt mit der Identifizierung elektronisch gespeicherter Informationen (Electronic Stored Information = ESI), die für die Angelegenheit relevant sein können, und endet mit der Erstellung von reaktionsschnellen, nicht privilegierten ESI für eine anfragende Partei. Dazwischen gibt es eine Reihe von Schritten, die darauf abzielen, Daten von der Identifizierung und Sammlung über die Verarbeitung, Überprüfung, Analyse und Produktion zu verschieben.

- *Die Schlüsselphasen für die Verarbeitungsphase des EDRM sind:*
  - *ESI-Aufnahme und Dateiextraktion*
  - *Anfängliche Filterung*
  - *Text-, Metadaten- und Bildextraktion*
  - *Ausgabe*
  - *Bericht*

Die Schlüsselprozesse sind in einem Reference Modell grafisch zusammengefasst:

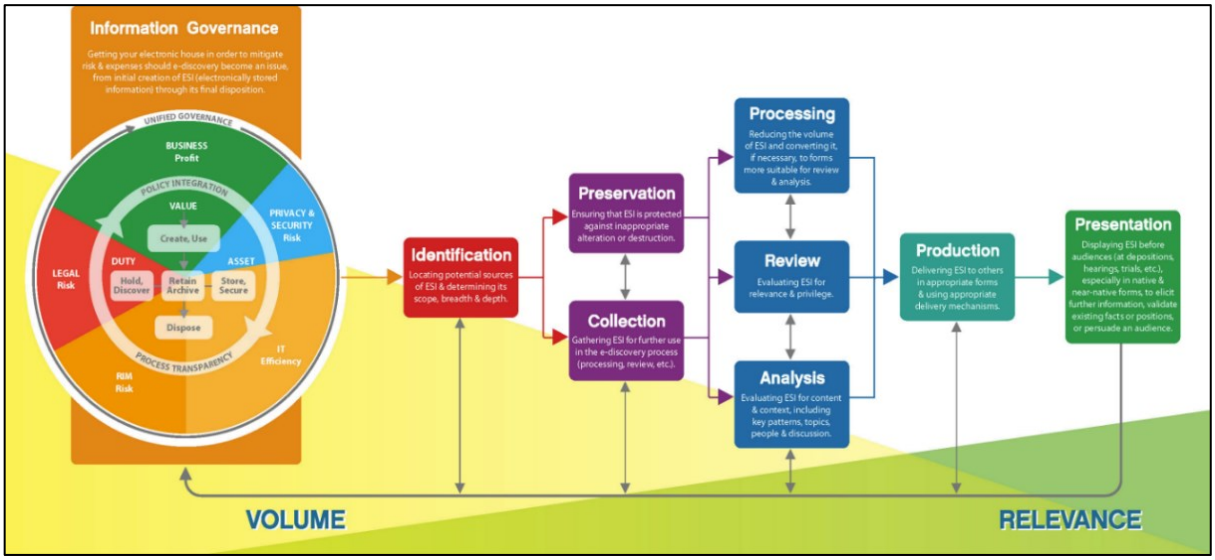


Abbildung 2.1: Electronic Discovery Reference Model

Ein deutschsprachiges Modell wurde von der Fa. FASTDETECT im Rahmen eines Vortrages zum 9. Fachanwaltstag IT-Recht am 19.10.2019 in München vorgestellt. Hier sind einzelne Prozessschritte aus dem Ursprungsmodell nocheinmal deutlicher in Phasen (Vorsortierung & Sichtung etc.) eingeteilt und mit allgemein verständlichen Bezeichnungen versehen worden, die für diese Arbeit nachfolgend genutzt werden.

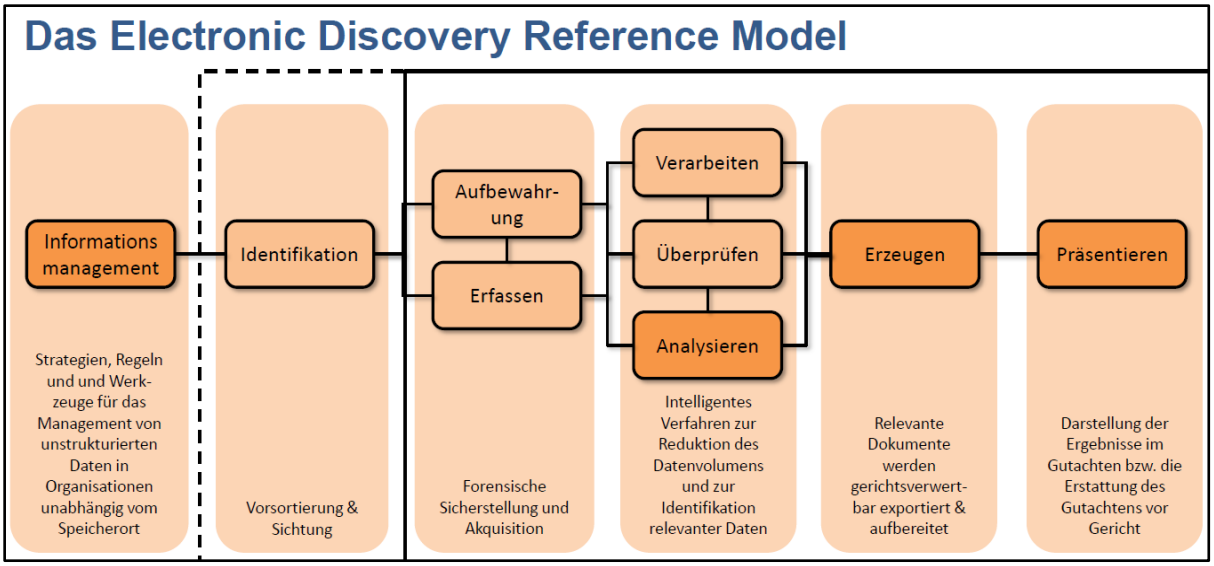


Abbildung 2.2 : Deutsche Interpretation des EDR-Modells

2.2 Information Governance (Informationsmanagement)

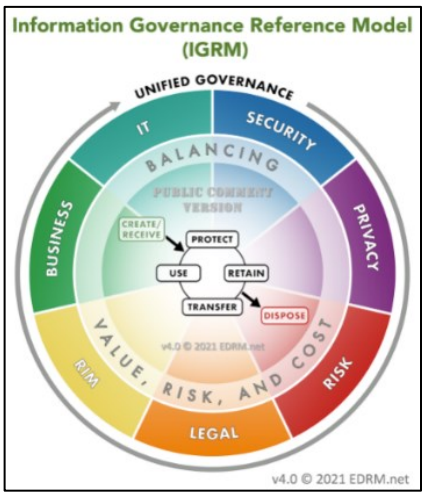


Abbildung 2.3: Information Governance Reference Model

Zu Beginn einer eDiscovery-Nutzung findet eine Aufnahme aller Beteiligten (Stakeholder [4]), der zu beachtenden Regeln und notwendigen Handlungen statt. Bei diesem Informationsmanagement geht es darum, *„...einen komplexen Satz interoperabler Prozesse zu konzipieren und die Verfahren und Strukturelemente in die Praxis umzusetzen..."* [5]. Hier werden neben der eigentlichen Datenerhebung, -verarbeitung und -auswertung auch alle anderen Aspekte wie Datenschutz, Aufbewahrung, Kosten, Risiken und Sicherheit mit einbezogen und daraus ein Gesamtkonzept bezogen auf die Umgebung der Stakeholder entwickelt. Hierfür wurde zusätzlich das Information Governance Reference Model (Abbildung 2.3) entwickelt. Die Beachtung dieser Regeln ist für die Vorbereitung eines IT-Projektes zur Einführung und Nutzung eines eDiscovery-Systems und auch für den späteren Betrieb wichtig und hilfreich, wird aber hier nicht näher betrachtet, da der Fokus dieser Arbeit auf den technischen Funktionalitäten, also dem Verarbeitungsprozess im Kontext einer polizeilichen Aufgabenstellung zur Aufbereitung und Darstellung von digitalen Beweismitteln liegt.

[4] Stakeholder = Interessengruppen (hier: Polizei u. Staatsanwaltschaft)
[5] https://edrm.net/resources/frameworks-and-standards/information-governance-reference-model/- The Outer Ring

2.3 Electronic Stored Information (ESI)

Bevor der Prozessablauf näher erläutert werden kann, ist das zu Grunde liegende digitale Verarbeitungsmaterial zu definieren. Elektronisch gespeicherte Informationen (= ESI) i. S. der US Federal Rules of Civil Procedure, Regel 34a

„...sind Informationen, die in digitaler Form erstellt, verändert, kommuniziert, gespeichert und am besten in digitaler Form genutzt werden und die die Verwendung von Computerhard- und –software erfordern...“ [6]

Diese Definition ist uneingeschränkt auch auf eDiscovery-Systeme für deutsche Strafverfahren übertragbar. Hierbei wird in drei verschiedene Datengruppen unterschieden, die verarbeitet werden können:

2.3.1 Arten von Datengruppen

Bei **strukturierten Daten** sind die Inhalte in voneinander abhängigen Feldern in Zeilen und Spalten geordnet. Derartige Daten finden sich in Datenbanken oder Tabellen.

Semistrukturierte Daten sind in Teilen geordnet; E-Mails beispielsweise haben eine Struktur für Angaben wie Absender und Adressat, der eigentliche Inhalt im Body jedoch ist aus Sicht einer Datenbank unbrauchbar.

Bei **unstrukturierten Daten**, auch Dark Data genannt, ist lediglich der Dateityp bekannt. Wie ihr Inhalt geordnet ist, kann nicht identifiziert werden. Unstrukturierte Daten entstehen im Nutzungsbetrieb durch den Anwender. Sie finden sich beispielsweise in Textnachrichten im Body oder als Anhang von E-Mail-Nachrichten. Sie kommen als Text-, Bild-, Audio- oder Videodateien sowie in anderen Dateiformaten vor.

2.3.2 Verfahrensrelevante Daten innerhalb unstrukturierter Massendaten

Dateien aus unstrukturierten Massendaten können in vielfältiger Art vorhanden sein. Grundsätzlich wird dabei aber in zwei Arten unterschieden:

- den Metadaten, die Informationen über die Dateiinhalte enthalten, also

[6] https://en.wikipedia.org/wiki/Electronically_stored_information_(Federal_Rules_of_Civil_Procedure)

a) wer ist der Autor eines Dokuments oder welche Auflösung, Blende, welcher Kameratyp wurden zur Bilderzeugung verwendet und

b) welchen Namen trägt diese Datei, welches Dateiformat wird verwendet, wer hat Zugriffsrechte und wann wurde dieses Image erzeugt, verändert oder zuletzt geöffnet

Beispiele:

a) EXIF (Exchangeable Image File Format)

Kameramodell	NIKON D200
Blendenzahl	F/8
Belichtungszeit	1/250 Sek.
ISO-Filmempfindlichkeit	ISO-200
Lichtwert	0 Schritt(e)
Brennweite	30 mm
Maximale Blende	4.1
Messmodus	Einpunkt
Abstand	
Blitzlichtmodus	Ohne Blitzlicht
Blitzlichtenergie	
35mm Brennweite	45

Abbildung 2.4: Metadaten EXIF-Information

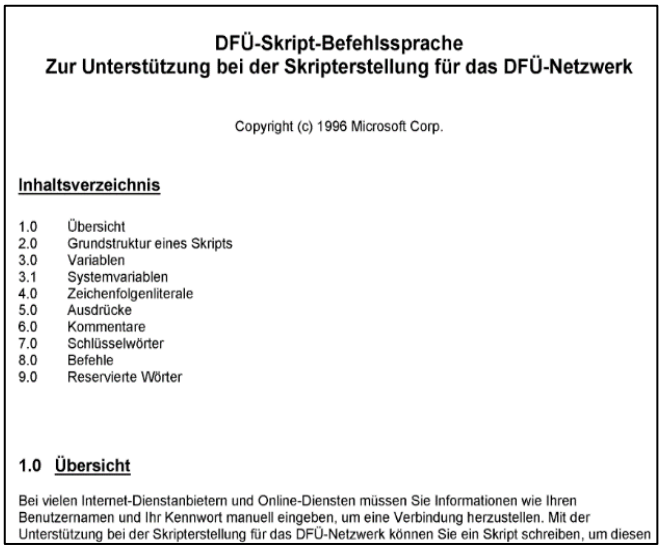
b) Timestamps

Letzter Zugriff	04/20/22 08:34:51
Datei erstellt	04/18/22 09:04:27
Zuletzt geschrieben	05/31/18 11:08:58
Eintrag geändert	04/18/22 09:09:36

Abbildung 2.5: Metadaten -Timestamp - Informationen

- den eigentlichen textbasierenden Inhaltsdaten, die die Informationen enthalten (also z.B. der Textinhalt eines Dokuments oder auch Tabelleninhaltsdaten

Beispiel:



DFÜ-Skript-Befehlssprache
Zur Unterstützung bei der Skripterstellung für das DFÜ-Netzwerk

Copyright (c) 1996 Microsoft Corp.

Inhaltsverzeichnis

1.0 Übersicht
2.0 Grundstruktur eines Skripts
3.0 Variablen
3.1 Systemvariablen
4.0 Zeichenfolgenliterale
5.0 Ausdrücke
6.0 Kommentare
7.0 Schlüsselwörter
8.0 Befehle
9.0 Reservierte Wörter

1.0 Übersicht

Bei vielen Internet-Dienstanbietern und Online-Diensten müssen Sie Informationen wie Ihren Benutzernamen und Ihr Kennwort manuell eingeben, um eine Verbindung herzustellen. Mit der Unterstützung bei der Skripterstellung für das DFÜ-Netzwerk können Sie ein Skript schreiben, um diesen

Abbildung 2.6: Textinhalte

Die Verarbeitungsschritte in einem eDiscovery-System lassen sich in drei Prozessgruppen einteilen:

- Identifizieren, Aufbewahren und Erfassen
- Überprüfen, Analysieren und Verarbeiten und
- Erzeugen (übertragen der Ergebnisse in die Datenbank(en)).

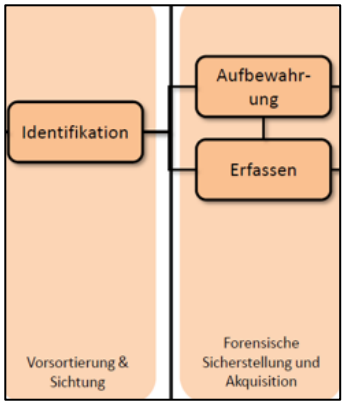
Der im EDRM vermerkte letzte Prozessschritt, das

- Präsentieren

bezieht sich auf empfohlene Darstellungsweisen vor Gericht und betrifft nicht mehr die technischen Anforderungen an ein eDiscovery-System. Sie sind jeweils individuelle Leistungen des Präsentierenden und keine technischen Anforderungen.

2.4 Identifikation, Aufbewahren und Erfassen

2.4.1 Identifikation



In dieser Phase werden die potentiellen ESI lokalisiert, die eine Relevanz für das spätere Strafverfahren haben könnten. Diese Stufe ist vergleichbar mit der Sicherstellung bzw. Beschlagnahme von digitalen Beweismitteln durch den polizeilichen Ermittler. Das bedeutet, der Stakeholder Ermittler entscheidet durch strafprozessuale Beweissicherungsmaßnahmen, welches der digitalen Medien eine mögliche Relevanz für das Strafverfahren hat und was somit im späteren eDiscovery-Prozess verarbeitet werden soll. Dieser Prozessschritt ist also noch keine eigentliche technische Funktion in einem eDiscovery-System, sondern eine manuelle Vorarbeit.

2.4.2 Aufbewahren und Erfassen

Hier finden die eigentliche Sicherung und Konservierung der digitalen Beweismittel statt. Dabei geht es darum, für das spätere Gerichtsverfahren die Originalität und Authentizität der digitalen Beweismittel nachzuweisen, so dass Löschungen oder Veränderungen nachträglich nicht mehr möglich sind. Dies wird üblicherweise durch das „Einfrieren“, also der Unveränderbarkeit der gesicherten Daten nach der Sicherung per Hashwert-Bildung erreicht. Auch dies ist eine Vorarbeit für die spätere Nutzung der Originaldaten im eDiscovery-Prozess.

In der IT-Forensik wird dieses „Einfrieren“ bzw. Sichern von relevanten Daten durch

- das physikalische Auslesen der gesamten digitalen Rohdaten aus den Speichermedien der Geräte (auch bekannt als physikalische oder forensische Sicherung) in Images oder durch

- technisch bedingte Teilsicherungen in Form von logischen Daten, die zum Beispiel bei der Auslesung eines E-Mail-Kontos eines Beschuldigten aus einem E-Mail-Server, Teildatensätzen aus Buchhaltungsdatenbanken oder einem Netzlaufwerk am Durchsuchungsort vorgenommen werden sowie

- der Erstellung von Prüfsummen über diese gesicherten Datenbestände

erreicht.

In der nachfolgenden Abbildung 2.7 ist diese Form der Sicherung und Hashwert-bildung beispielhaft dargestellt:

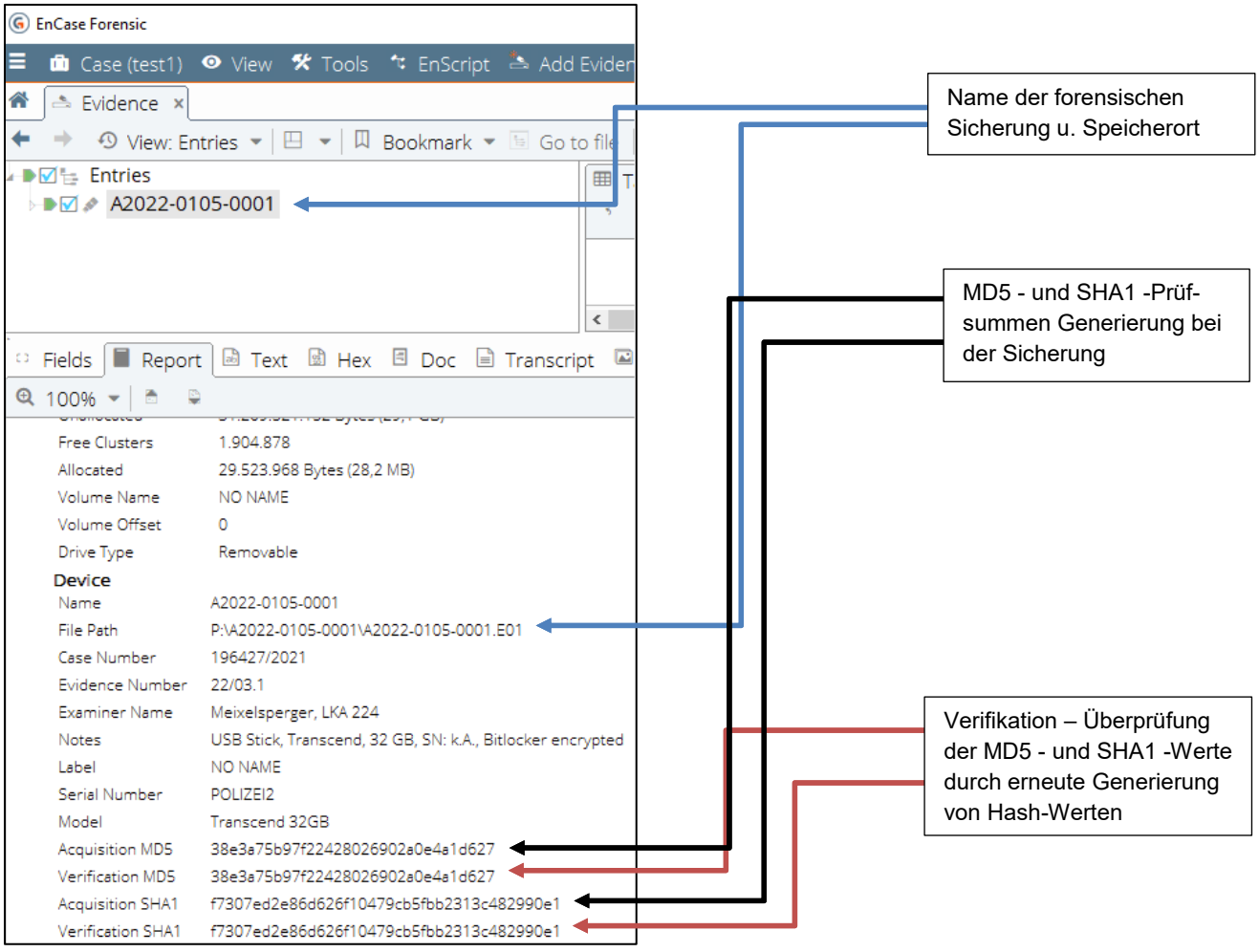
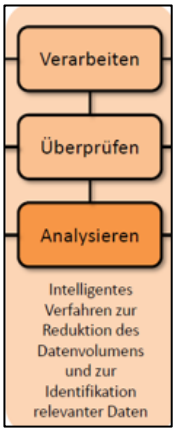


Abbildung 2.7: Beispieldarstellug einer Sicherung und Verifizierung unter Encase Forensic 8

2.5 Verarbeiten, Überprüfen und Analysieren



Nachdem die manuellen Vorarbeiten abgeschlossen sind, geht es in die eigentliche Verarbeitung der Daten. In diesem zweiten Prozessschritt finden die

- Datenreduktion
- Datenextraktion, also Freilegung von Sicherungs- und Komprimierformaten
- Überprüfung und Erkennung auf Relevanz (Review) von Datentypen sowie
- deren Analyse

statt.

2.5.1 Datenreduktion vor und nach dem Verarbeitungsprozess

Die Verarbeitung von Originaldaten in eDiscovery-Systemen kann sehr lange dauern. Dies ist abhängig von

- der Programmierung der Verarbeitungs-Engine eines eDiscovery-Systems, also welchen Umfang die jeweiligen Prüf- und Verarbeitungsalgorithmen umfassen
- der eingesetzten Hardware (Anzahl der Prozessoren, RAM-Größe, Durchsatz im Datenbus, Speichermedium etc.)
- der Art des Lizenzumfangs des verwendeten eDiscovery-Produkts (im NUIX Lizenzmodell wird mit Workern gearbeitet - je mehr Worker umso höher die Verarbeitungsleistung)
- dem Umfang der zu verarbeitenden Datenmenge und
- den Konfigurationseinstellungen zum Verarbeitungsprozess (Processing)

Viel Verarbeitungszeit kann dadurch gespart werden, indem der Originaldatenbestand, also die gesicherten Daten, vor der eigentlichen Verarbeitung und auch danach auf das reduziert werden, was wirkliche Verfahrensrelevanz besitzt, indem alle anderen nicht benötigten Daten aussortiert oder gar nicht erst verarbeitet werden.
Der Oberbegriff für das Reduzieren mit verschiedenen Techniken lautet hierfür Culling. Im Allgemeinen werden vier Arten von Culling unterschieden: DeNISTing, Datei (MIME)-Typ-Filterung, Deduplizierung und Predictive Coding.

DeNISTing

Der Begriff DeNISTing ist ein Kunstwort und soll die Nutzung von Hashes aus der National Software Reference Library (NSRL) des National Institute of Standards and Technology der USA (NIST) zur Entfernung von nicht relevanten Dateien beschreiben. Die Library wird immer wieder aktualisiert und enthält jeweils Hashes der Programm- und Systemdateien bekannter Softwareanbieter wie z.B. Microsoft, Adobe aber auch Red Hat oder Apple´s macOS. Der Einsatz ist sowohl vor, als auch während der eDiscovery-Nutzung möglich, indem man

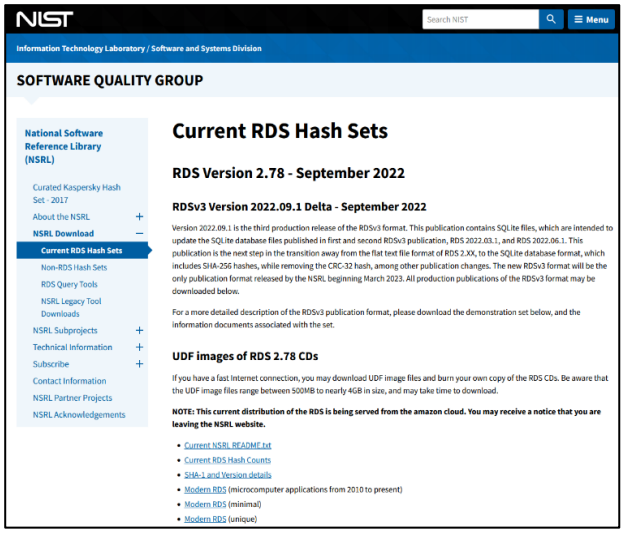


Abbildung 2.8: White Hashes von der NSRL

a) ein klassisches IT-Forensik-Tool für die erstellten Datensicherungen einsetzt, alle enthaltenen Dateien „hasht“ und über die im Tool importierte Hash-Liste der NIST derartige nicht relevante Dateien durch Vergleich identifiziert und vom gesicherten Datenbestand abtrennt bzw. wegfiltert oder

b) indem man in der eDiscovery-Anwendung selber die NIST-Liste *“…in Verbindung mit den Dateisignaturen i.d.R. mit eDiscovery-Programmdatenbanken verwendet, um Dateisignaturen gesammelter Daten zu Ermittlungszwecken zu vergleichen. Jede Datei, die mit einer Datei in der NIST-Liste übereinstimmt, wird “de-NISTed” – das heißt, ausgeschlossen und nicht weiter verarbeitet oder analysiert.”* [7]

Die so reduzierte Verarbeitungsgröße verbraucht dann weniger Rechenzeit und ist somit schneller fertig gestellt.

MIME-Type-Filterung

EDiscovery-Systeme fassen Dateitypen wie z.B. aus dem Office- (pptx, wps, docx etc.) oder auch Bild-Bereich (jpg, bmp, png etc.) in Gruppen zusammen. Für diese Gruppen themengleicher Dateitypen wird der Begriff „Multipurpose Internet Mail Extensions“

[7] https://edrm.net/resources/frameworks-and-standards/edrm-model/edrm-stages-standards/edrm-processing-standards-guide-version-1/ - Verwendung der NIST Liste bei der e-Discovery Verarbeitung

(MIME) verwendet, der ursprünglich eine Technik aus der Kommunikation zwischen Webserver und Browser beschreibt. Der MIME-Typ gibt an, um welche Art bzw. Klasse es sich bei den gesendeten Daten handelt.
Durch eine Konfiguration dieser MIME-Typen in der Verarbeitungs-Engine wird festgelegt, was innerhalb der untersuchten Datenmenge Relevanz hat oder nicht. Mit den jeweiligen Konfigurationen kann dann eine Verarbeitungs-Engine bestimmte Dateitypen bzw. -gruppen auslassen und die zu analysierenden Rohdaten werden dadurch schneller verarbeitet.

Im ersten Beispiel (Abbildung 2.9, rot markiert), lassen sich so im eDiscovery-System NUIX Workstation alle Überprüfungen von Betriebssystemdateien und Protokollen auslassen, wenn sie für ein Verfahren keine Relevanz besitzen:

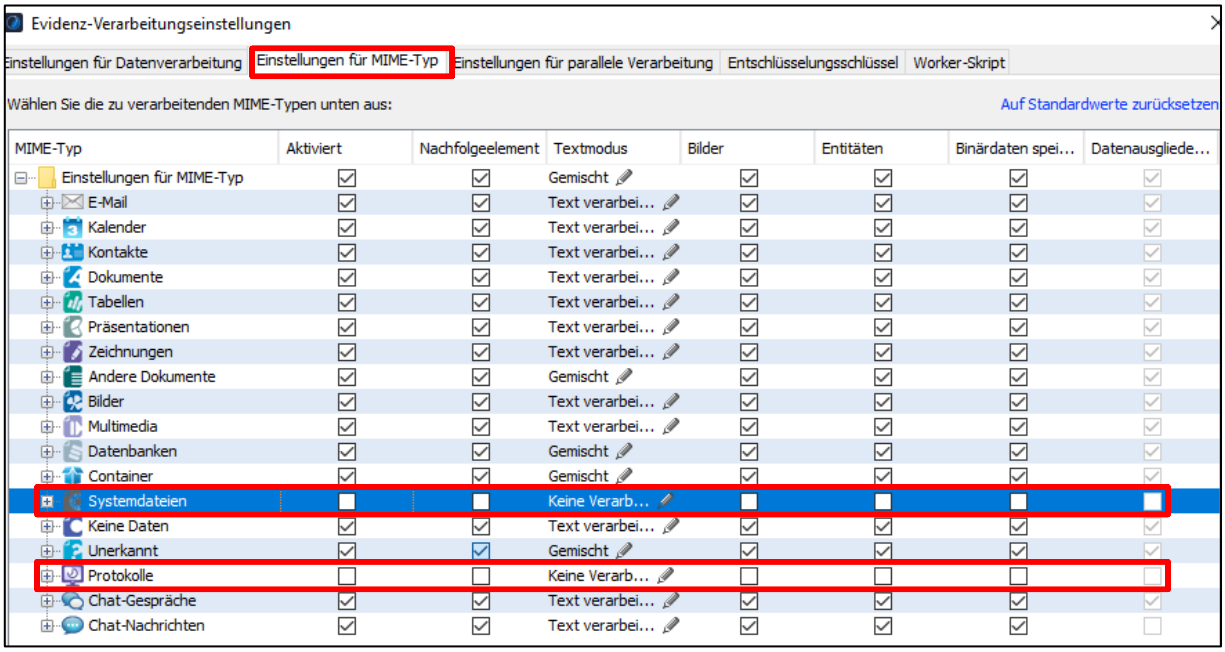


Abbildung 2.9 Processing-Einstellung bei NUIX WS 8.8

Im zweiten Beispiel (Abbildung 2.10, rot markiert), der eDisocvery-Anwendung AXIOM Process können z.B. diejenigen Filter zur Verarbeitung von Smartphone-Backups deaktiviert werden, wenn die zugrundeliegende Datensicherung von einem Arbeits-platz-PC stammt, der keine Backups von persönlichen Smartphones zulässt:

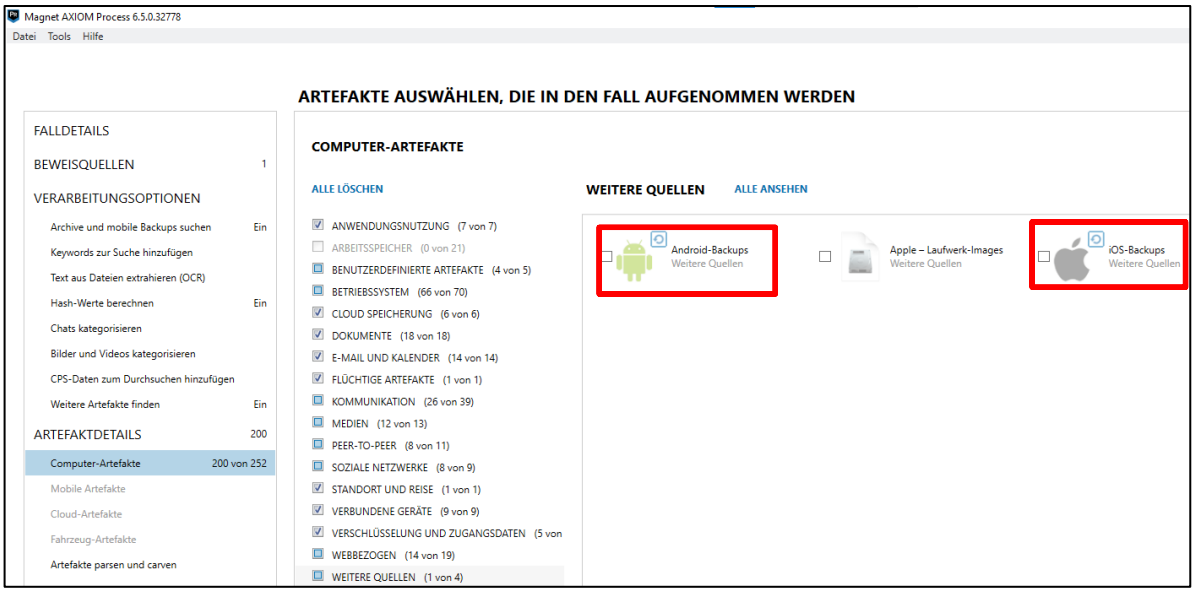


Abbildung 2.10: Processing-Einstellung bei AXIOM 6.8

Deduplication (Deduplizierung)

Häufig weisen die zu verarbeitenden Daten identische Kopien einzelner Dateien auf. In einem NUIX-Block-Beitrag [8] berichtet der Autor Corey Tomlinson, dass nach seinen Erfahrungen die Dubletten 30 – 50% Anteil am Gesamtdatenvolumen haben. Deduplizieren kann sowohl vor (mit anderen Forensiktools) als auch nach dem Verarbeitungsprozess innerhalb der eDiscovery-Anwendung durchgeführt werden. Dann wird allerdings keine Verarbeitungszeit eingespart, da die Prüfsummenbildung teil des Verarbeitungsprozesses ist.

In den Richtlinien für die eDiscovery-Verarbeitung wird im Dokument *2.0 Erstfilterung* [9] unter *2.2* auf *„Doppelte Dateien identifizieren und entfernen“* hingewiesen. In Abbildung 2.11 ist diese Funktionalität unter NUIX Workstation 8 zu sehen (rot markiert).

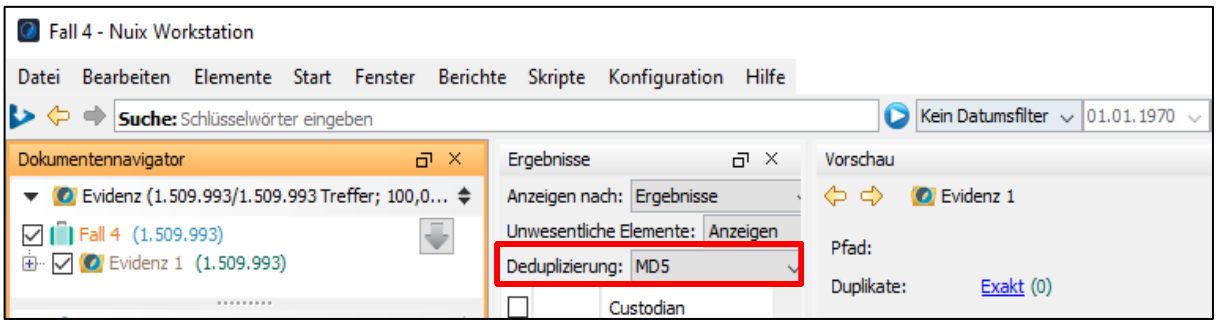


Abbildung 2.11: Deduplizierung unter NUIX

[8] https://www.nuix.com/blog/tying-deduplication-your-ediscovery-bottom-line
[9] https://edrm.net/wiki/2-0-initial-filtering/

Predictive Coding

Dieser noch neue Ansatz einer Datenreduktion verwendet

„...intelligente, maschinelle Lernfunktionen, mit denen man große Mengen an Fall-inhalten aussortieren kann, die für ihre Untersuchung nicht relevant sind. Dies wird erreicht, indem man eigene prädik-tive Codierungsmodelle erstellt und trainiert, mit denen die relevantesten Elemente für die Überprüfung priorisiert werden. Für den Anfang wird ein Modell erstellt indem nur 50 Elemente als relevant oder nicht relevant gekennzeichnet werden. Das System verwendet dann dieses Training, um Vorher-sageergebnisse auf jedes Element im Überprüfungssatz anzuwenden. Auf diese Weise können Elemente basierend auf der Vorhersagebewertung gefiltert werden, sodass man zuerst die relevantesten (oder nicht relevanten) Elemente überprüfen kann. “ [10]

Diese Technologie ist ein Lösungsansatz, bei dem automatisiert über generierte Filter-modelle vollautomatisch im Gesamtbestand gesucht werden kann. Durch fortgesetz-te Verfeinerung (Trainierung) der Filtermodelle werden die Filterergebnisse immer weiter verbessert. Damit wird deutlich, dass hier Künstliche Intelligenz (KI)-Technolo-gie für die Datenreduktion bzw. -filterung eingesetzt wird. Ein EDRM-Leitfaden oder eine Richtlinie gibt es hierzu noch nicht. Das Thema wird in Block-Beiträgen in der EDRM-Community aber bereits behandelt [11].

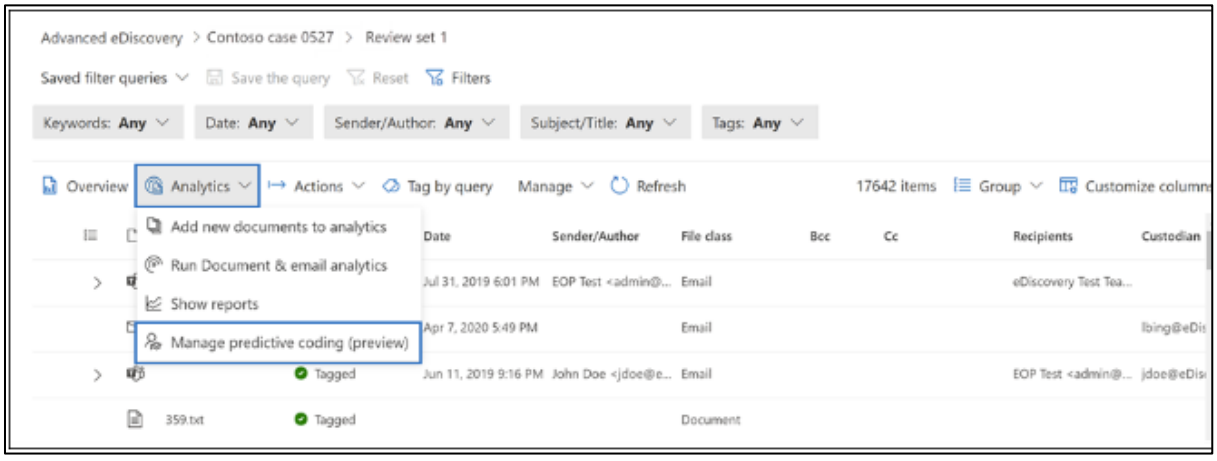


Abbildung 2.12: Predictive Coding bei Microsoft´s eDiscovery Anwendung

Fazit

Das Culling, also das Wegfiltern von Daten, spielt in der klassischen IT-Forensik eine nachgeordnete Rolle, denn erst durch die Untersuchung dieser Daten durch einen IT-

[10] https://learn.microsoft.com/en-us/microsoft-365/compliance/predictive-coding-quick-start?view=o365-worldwide
[11] https://edrm.net/2022/06/what-is-predictive-coding-in-ediscovery/

Forensiker wird entschieden ob und welche Filtermöglichkeiten (z.B. DeNisting) angewendet werden. Ansonsten besteht die Gefahr, dass mögliche verfahrensrelevante Spuren von vorne herein durch Filtermaßnahmen ausgeschlossen und so nicht mehr zur Bewertung durch einen IT-Forensiker gelangen.
EDiscovery-Anwendungen fokussieren auf Nutzerdaten und selten auf bekannte System- und Programmdateien. Ihr Ziel ist es, Kosten zu sparen. Je weniger Daten verarbeitet und ausgewertet werden müssen, umso mehr begründet es seinen Einsatz.

2.5.2 Datenextraktion

Neben logischen Sicherungen können je nach verwendetem eDiscovery-System auch forensische Sicherungen oder die Images von virtuellen Maschinen (VM) verarbeitet werden. Um die relevanten Daten zu erkennen, müssen eDiscovery-Systeme in der Lage sein, die unterschiedlichen Format-Schichten zu erkennen, zu öffnen bzw. zu extrahieren und die enthaltenen ESI dem Verarbeitungsprozess zuzuführen. Im nachfolgenden Workflow ist dies beispielhaft dargestellt:

Schicht 1 (Image Formate von forensischen Sicherungen oder VMs)
Forensische Image-Formate (*.E01, *.L01, *.DD, *.AFF, *.CTR usw.) enthalten nicht nur einzelne Dateien oder Verzeichnisse, sondern Abbilder von ganzen Partitionen, Laufwerken oder
Images von virtuellen Maschinen (z.B. *.vmdk, *.vhd bzw. *.vhdx, *.vdi etc.)

Schicht 2 (Dateiverwaltungssysteme)
Dateiverwaltungsformate (z.B. NTFS, HPFS+, EXT3, EXT4, FAT16, FAT32, exFAT)

Schicht 3 (Verzeichnisstruktur des jeweils verwendeten Betriebssystems)
Betriebssysteme (z.B. Windows XP, 7, 8, 10, Ubuntu, Debian, Unix, macOS)

Schicht 4 (komprimierte und E-Mail Container-Formate)
In dieser Schicht werden Daten aus gängigen Dateitransportcontainerformaten wie ZIP

oder RAR entkomprimiert und die darin enthaltenen eigentlichen Dateien auf ihre mögliche Verfahrensrelevanz anhand des Formats überprüft und ggf. einem weiteren Verarbeitungsprozess zugeführt. Genauso verhält es sich mit E-Mail-Containern aus Anwendungen wie MS Outlook oder IBM Lotus Notes und den dort in Textnachrichten eingebetteten Objekten (Attachements). Werden ihre Dateiformate (z.B. pst, docx, jpg, usw.) vom eDiscovery-System erkannt, so werden diese Dateien dem Verarbeitungsprozess zugeführt.

Verdeutlicht wird dieses „Extrahieren“ unterschiedlicher Schichten und damit dem Detektieren relevanter Daten in der nachfolgenden Abbildung 2.13:

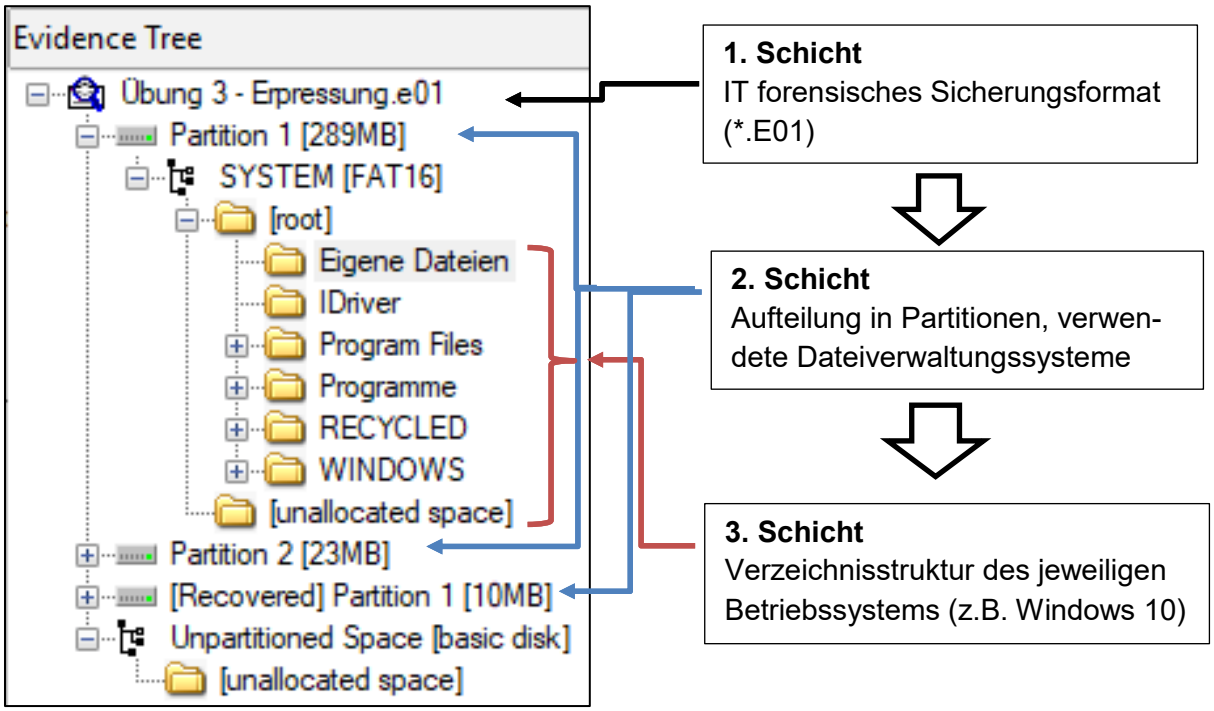


Abbildung 2.13: Darstellung der Schichten unter FTK Imager 4

Aus der nachfolgenden Abbildung 2.14 wird deutlich, was die Programmierung eines eDiscovery-Systems leisten muss, um nach dem Erkennen der Aufteilung des Datenträgers (Partitionierung) in der 2. Schicht das eingesetzte Dateiverwaltungssystem und „aufgesetzte“ Betriebssystem zu erkennen und damit den entsprechenden Programmcode eines eDiscovery-Programms „aufzurufen“, der auf der Basis dieser Verzeichnisstruktur die darunterliegenden ESI in der 3. und 4. Schicht erkennt und enthaltene mögliche verfahrensrelevante Dateien herausfiltern kann:

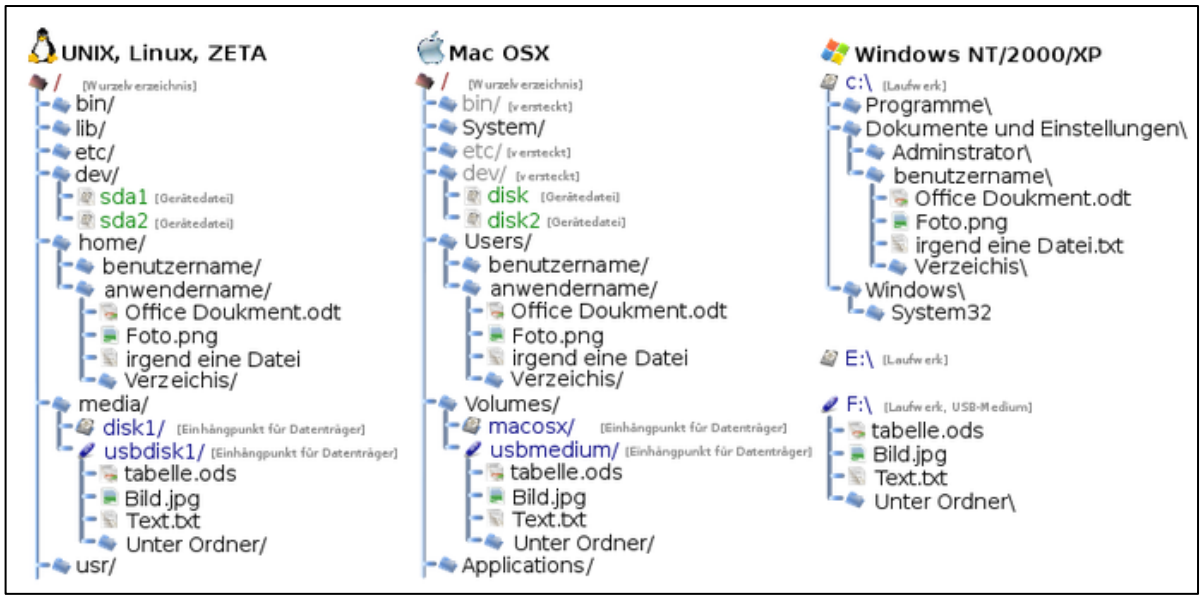


Abbildung 2.14: Verzeichnisstrukturen in verschiedenen Betriebssystemen

2.5.2.1 Freilegung der Schichten

1. Schicht - Forensische oder Virtual Machine Images

Damit ein Datenbestand ausgelesen werden kann, muss zuerst bekannt sein, wo innerhalb einer Datensicherung die eigentlich zu analysierenden Originaldaten beginnen und enden. In der 1. Schicht muss somit geklärt werden, um was für ein forensisches Sicherungs- oder virtuelles Maschinen-Format es sich handelt und wo innerhalb dieses Images die tatsächlich gesicherten Daten beginnen und enden. Zusätzlich zerteilen forensische Sicherungsprogramme wie auch virtuelle Maschinen-Programme die Gesamtmenge gesicherter Daten in kleinere Segmente, da sich so die Abspeicherung auf mehreren Datenträgern (in den Anfängen z.B. auf DVDs) und der Versand an Dritte einfacher umsetzen lässt.

Am nachfolgenden Beispiel (Abbildung 2.15) wird für das weitverbreitete Expert Witness Format (besser bekannt als E01) der Analyseprozess eines eDiscovery-Systems bei einem zu verarbeitenden forensischen Image verdeutlicht, das in die erwähnten Segmente (Section files) unterteilt ist:

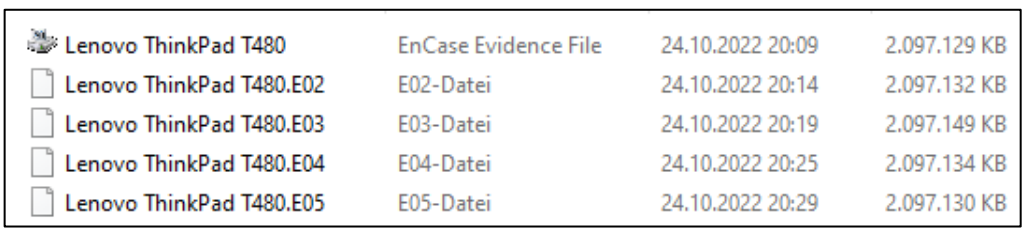


Abbildung 2.15: Forensische Sicherung aufgeteilt in Section Files

Jeder dieser Section-Files (*.E0x) hat seinen eigenen Header, der sich je nach verwendetem Sicherungsformat unter scheidet. Auf Github [12] findet sich für das meistverbreitete forensische Sicherungsformat EWF (Expert Witness Format) eine Header-Definition (siehe Abbildungen 2.16 und 2.17):

The file header is 13 bytes of size and consists

Offset	Size	Value	Description
0	8		Signature "EVF\x09\x0d\x0a\xff\x00"
8	1	0x01	Start of fields
9	2		Segment number Must be 1 or higher
11	2	0x0000	End of fields

Abbildung 2.16: File Header Information für EWF-Dateien

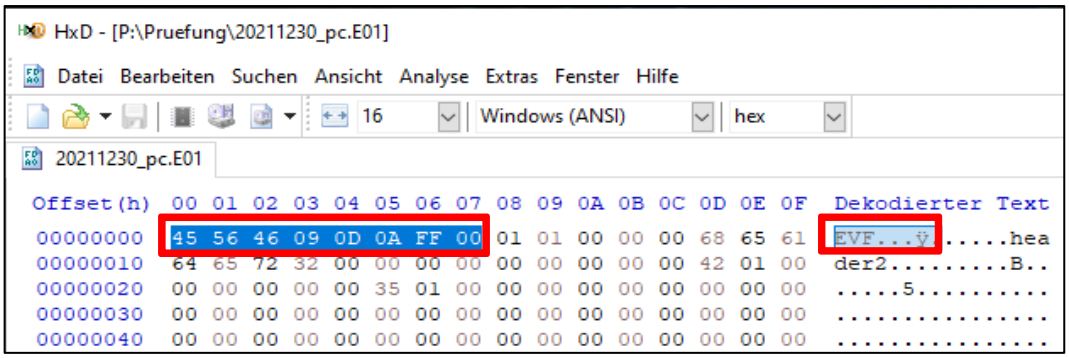


Abbildung 2.17: EWF-Signatur als HEX-Werte

Als nächstes muss das eDiscovery-Tool identifizieren, wo innerhalb dieses File-streams die tatsächlich gesicherten Rohdaten nach dem EWF-Header beginnen. Da forensische Images platzsparend erstellt werden, wird zusätzlich ein Komprimierformat genutzt. Für das EWF-Format ist dies Zlib (siehe Abbildung 2.18):

5. Kompression

5.1. Zlib-komprimierte Daten

Die komprimierten Daten werden im komprimierten Datenformat zlib (RFC1950) gespeichert. Dieses Format verwendet Big-Endian.

Die komprimierten Daten sind variabel und bestehen aus:

Offset	Größe	Wert	Beschreibung
0.0	4 bit		Komprimierungsmethode
0.4	4 bit		Informationen zur Komprimierung
1.0	5 bit		Prüfen Sie Bits
1.5	1 bit		Voreingestelltes Wörterbuch-Flag
1.6	2 bit		Kompressionsstufe

Abbildung 2.18: Komprimierformat für EWF

[12] Wikipedia: GitHub ist ein netzbasierter Dienst zur Versionsverwaltung für Software-Entwicklungsprojekte. Namensgebend war das Versionsverwaltungssystem Git. Das Unternehmen GitHub, Inc. hat seinen Sitz in San Francisco in den USA. Seit dem 26. Dezember 2018 gehört das Unternehmen zu Microsoft.

Ist der Beginn des Sicherungsimages gefunden, müssen diese komprimierten Teilrohdaten wieder zu einem ganzen Image zusammengefügt werden, damit der Master Boot Record (MBR) oder die GUID Partitionstabelle (GPT) analysiert und die jeweils identifizierten Partitionen eingehängt (gemountet) werden können.
In der Abbildung 2.19 ist dann dieser Zustand mittels des Forensik-Tools Encase Forensics 8 mit einem MBR formatierten Datenträger beispielhaft dargestellt:

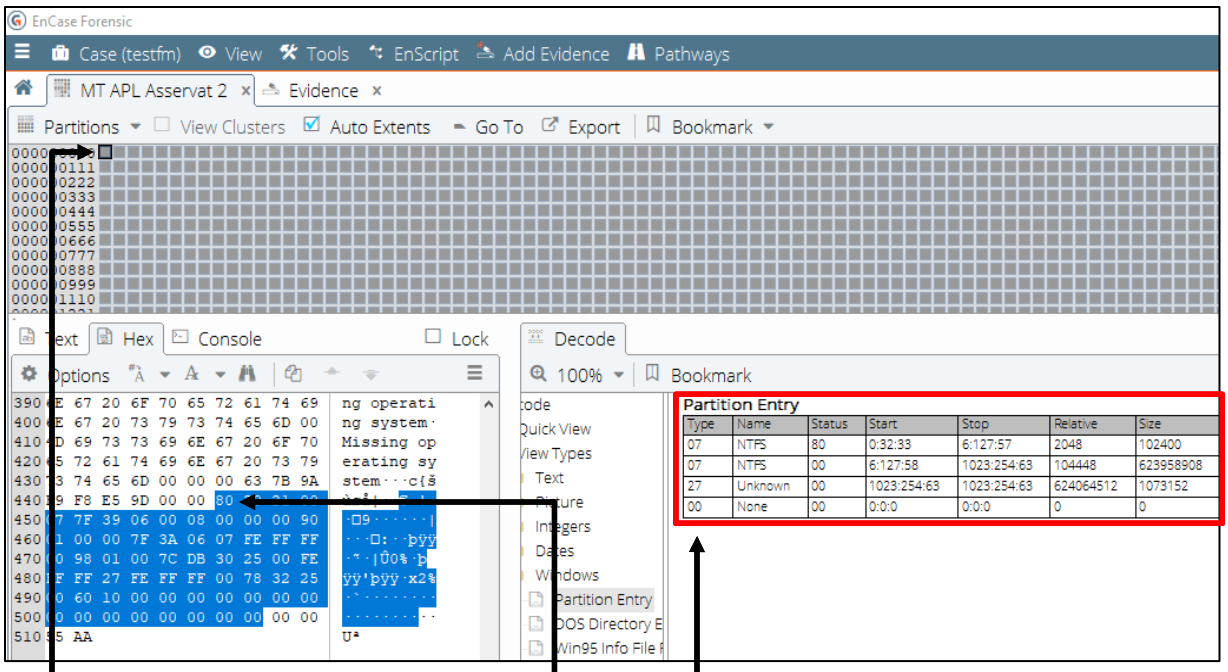


Abbildung 2.19: Master Boot Record mit Partitionstabelle

1. Sector im gesicherten Datenbestand: Master Boot Record mit enthaltener (blau markierter) Partitionstabelle ab Offset 446

Analyseergebnis der vier 16 Bit großen Partitionsinformationen inklusive Anfang und Ende des jeweiligen Volumes

Die am Anfang jedes Section Files stehenden Headerdaten sind:

- herausgefiltert bzw. abgetrennt
- der Beginn und das Ende der tatsächlich gesicherten Daten identifiziert und so
- können die Teilimages wieder zu einem Gesamtimage zusammengefügt werden,

so dass der nächste Analyseschritt - die Auslesung der Partitionstabelle - und somit das „Einhängen“ der Volumes möglich wird.

2. Schicht – Partitionierung, Aufteilung der Datenbereiche

Nun kann im nächsten Schritt angefangen werden, diese Volumes (Partitionen) anhand der Offset-Informationen aus der Partitionstabelle (siehe rote Markierungen in Abbildung 2.19) zu mounten (einzuhängen). In diesem Schritt analysiert ein eDiscovery – System den 16 Bit großen Informationseintrag in der jeweiligen Partition (Volume Boot Record = VBR), um so zu erkennen, welches Dateiverwaltungssystem (NTFS, exFAT, FAT32, EXT3, HPFS etc.) für dieses Volume eingesetzt wird.

In der Abbildung 2.20 wird am Sector 2048 (Information aus der Partitionstabelle) begonnen, den Dateninhalt einer Partition auszulesen und zu analysieren. Bei Dateiverwaltungssystemen wie FAT32, exFAT und NTFS wird dieser Partitions-Startbereich als Volume Boot Record (VBR) bezeichnet. Hier beginnt die 8 Byte lange Information des verwendeten Dateiverwaltungssystems NTFS im 4. Byte des 1. Sektors der Partition (siehe rote Markierungen):

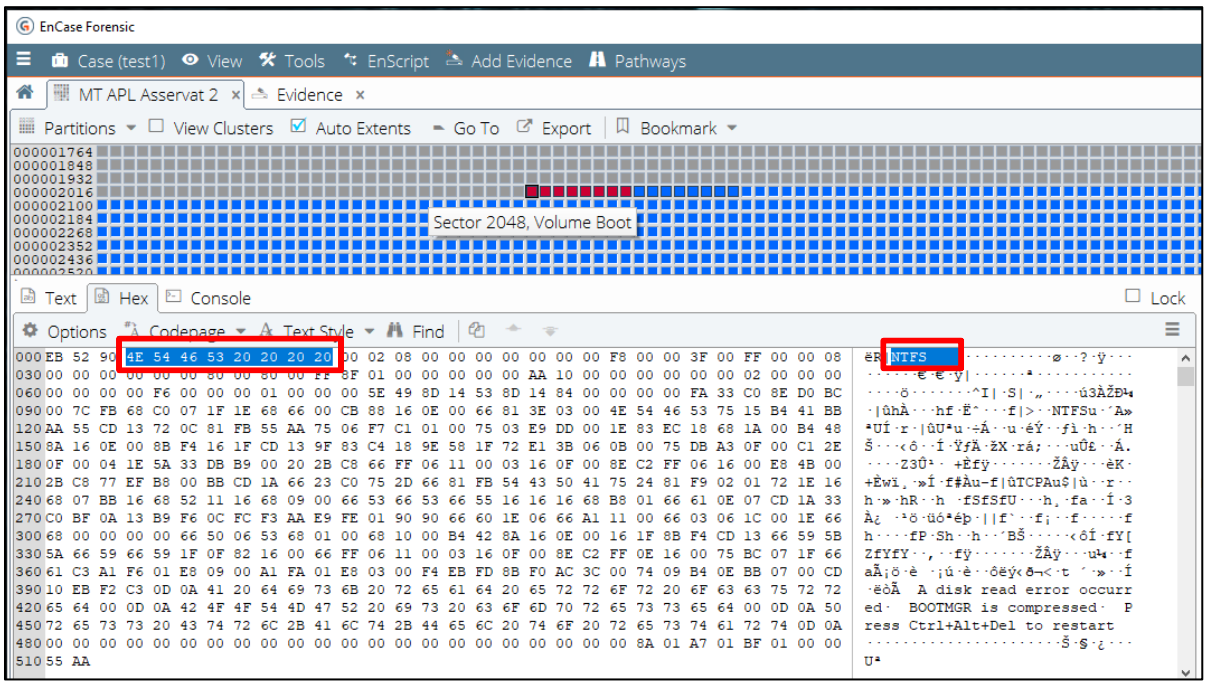


Abbildung 2.20: Volume Boot Record – Startbereich einer Partition

Bei einem Linux Dateiverwaltungssystem werden andere Partitionstypen wie Swap, Native oder Logical Volume (LVM) verwendet, die sich in Aufbau und Struktur im Vergleich zu Microsoft-basierenden Volume Types grundlegend unterscheiden. Eine weitere Variante zu den hier vorgestellten Dateiverwaltungssystemen bietet das von Apple entwickelte Dateiverwaltungssystem namens Apple File System (AFS) als Nachfolger von HFS+.

3. Schicht - Dateiverwaltungssystem

Im letzten Schritt kann das eDiscovery-Programm darauf basierend

- das verwendete Dateiverwaltungssystem,
- deren Verzeichnisstruktur und
- enthaltene verfahrensrelevante Dateien abbilden bzw. sichtbar machen

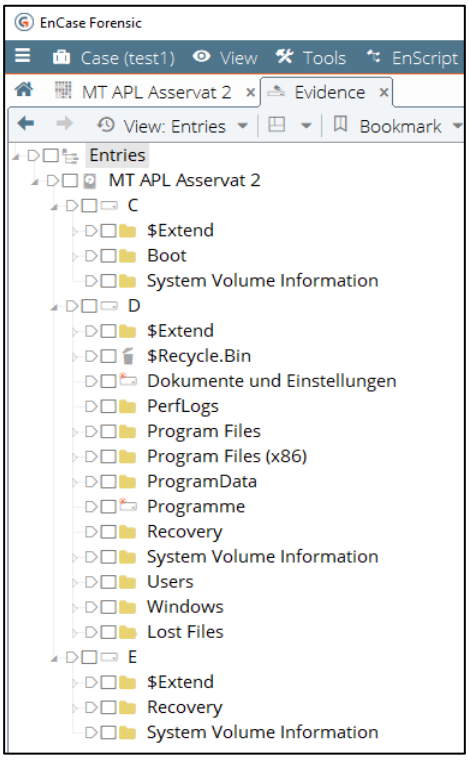


Abbildung 2.21:
Verzeichnisstruktur unter Windows

Damit ist der Teilprozess der „Freilegung der Daten“, die Datenextraktion eines Sicherungsimages durch

- Erkennen der Partitions-Aufteilung,
- dem jeweiligen Mounten der identifizierten Volumes und
- dessen verwendetem Dateiverwaltungssystem

abgeschlossen.

Fazit

Um die beschriebenen Prozessschritte über eine Programmierung zu automatisieren und zu realisieren, wird deutlich, dass dies auf unveränderbaren Gesetzmäßigkeiten, also der immer gleichen Struktur der Headerinformationen, der Lage der erwarteten Daten an bestimmten Stellen im Offset der Formatierung und des verwendeten Dateiverwaltungssystems beruht. Eine Programmierung erwartet, dass die gleichen Komprimierformate (hier im Beispiel Zlib) angewendet werden, um so den richtigen Punkt für den Beginn der tatsächlich gesicherten Daten berechnen zu können.
Bis zu diesem Verarbeitungsschritt unterscheiden sich eDiscovery-Systeme für IT–forensische Zwecke in keiner Weise von klassischen IT-Forensik-Tools, denn nur so sind die nächsten Verarbeitungsschritte überhaupt erst möglich.

2.5.3 Identifizierung und Überprüfung von Dateitypen

Nachdem durch Extraktionen die zu verarbeitenden Daten offengelegt wurden, geht es nun darum, die Dateitypen zu identifizieren, für die das eDiscovery-System eine Verarbeitungslösung vorhält. Im EDRM-Referenzdokument *1.0 ESI-Aufnahme und Dateiextraktion* heißt es unter 1.3 [13]:

„Es ist zwingend erforderlich, dass das Verarbeitungssystem die verschiedenen Dateitypen erkennt, die zur Aufnahme empfangen werden. Office-Dateien müssen ordnungsgemäß identifiziert werden; Bilddateien müssen als Bilder behandelt werden. Audio- und Video-Dateien müssen ebenfalls ordnungsgemäß adressiert werden. Eine falsche Identifizierung von Dateitypen kann schnell zu Verarbeitungsfehlern führen...“

und

„...Um Dateitypen zu identifizieren sollte die Verarbeitungssoftware in der Lage sein, eine korrekte Identifizierung von Dateitypen basierend auf mehreren Faktoren, einschließlich Header-Informationen, MIME-Typen und Dateierweiterungen vornehmen zu können.“

In der Abbildung 2.22 wird beispielhaft dargestellt, welche Dateitypen das Microsoft eigene eDiscovery-System namens *Purview* unterstützt:

Abbildung 2.22: Unterstützte Dateitypen unter MS Purview

Identifizierung über Dateinamenserweiterungen

Ein Dateiname besteht aus zwei Teilen: dem eigentlichen Namen und einer aus drei oder vier Buchstaben bzw. Zahlen bestehenden Endung, die durch einen Punkt getrennt sind (Beispiel: Video.mp4 oder auch Brief.docx). Diese „file extension“, auch genannt Datei-Suffix, wird für die Identifizierung relevanter bzw. unterstützter Dateiformate genutzt.

[13] https://edrm.net/wiki/1-0-esi-ingestion-and-file-extraction/

Identifizierung von komprimierten Dateien und E-Mail-Containern

Eine spezielle Gruppe bilden komprimierte Dateien und E-Mail-Container, denn sie beinhalten innerhalb ihrer Datei weitere Verzeichnisstrukturen. Dazu gehören

- Dateien von Komprimierprogrammen wie z.B. ZIP, RAR oder 7ZIP und
- Dateien aus (Internet) E-Mail-Systemen wie z.B. PST, NSF, OST, EDB, MSG

Bei diesen Formaten müssen die eDiscovery-Systeme in der Lage sein, die in diesen Containern vorhandenen Inhaltsdateien und Metadaten zu dekomprimieren und die Verzeichnisstrukturen für den Identifizierungsprozess, um welche Dateitypen es sich innerhalb der Container handelt, offenzulegen. Das Referenzdokument *1.0 ESI-Aufnahme und Dateiextraktion* [14] merkt hierzu unter 1.2 an:

„E-Mail-Sammlungen werden in der Regel in einer Containerdatei transportiert, die als PST (Personal Storage Table) oder OST für lokale temporäre Kopien und NSF für Lotus Notes bezeichnet wird. PSTs werden erstellt, um Microsoft Exchange-Dateien wie Outlook-E-Mails, Kontakte, Aufgaben und Kalenderelemente zu speichern. Um diese Informationen ordnungsgemäß zu extrahieren, muss das Verarbeitungssystem das entsprechende Codierungsschema anwenden, um Nachrichten und andere einzelne Elemente zu trennen, um ihren Inhalt erfolgreich zu extrahieren.“

Identifizierung und Klassifizierung über eine File Signatur Analyse

Eine falsche Identifizierung von Dateitypen kann schnell zu Verarbeitungsfehlern führen, da es nicht mit der dafür vorgefertigten Lösungsmatrix geöffnet und damit die enthaltenen Meta- u. Inhaltsdaten korrekt verarbeiten bzw. indexieren kann. Ohne den Einsatz von File-Type-Analysen im Verarbeitungsprozess wäre es sonst Tätern möglich, durch automatisiertes Umbenennen von Dateiendungen (bekannte Tathandlungen im Bereich Kinderpornographie oder Hacking) mittels Einsatz von Tools automatisch ganze Gruppen von Dateiendungen dieser relevanten Dateien einem Verarbeitungsprozess eines eDiscovery-Systems zu entziehen, da diese nicht erkannt und damit verarbeitet werden können.

Die Information über einen Dateityp wird typischerweise von einer Dateisignatur begleitet, die oft als „magic number“ bezeichnet wird. Eine Dateisignatur ist in der Regel 1 – 4 Byte lang und befindet sich bei der Überprüfung von Rohdaten beginnend im Offset 0 des Filestreams einer Datei. Darüber ist es dem Betriebssystem möglich,

[14] https://edrm.net/wiki/1-0-esi-ingestion-and-file-extraction/

das richtige Programm zum Lesen der Inhalts- und Metadaten zur Verfügung zu stellen.

Bestimmte Dateien haben Dateisignaturen, die über die genannte 4 Byte Länge hinausgehen oder nicht im Offset 0, sondern irgendwo innerhalb des Filestreams beginnen. In der nächsten Abbildung 2.23 ist am Beispiel eines PDF-Dokuments im Offset 0 beginnend die Dateisignatur in HEX- und ASCII-Code-Schreibweise dargestellt:

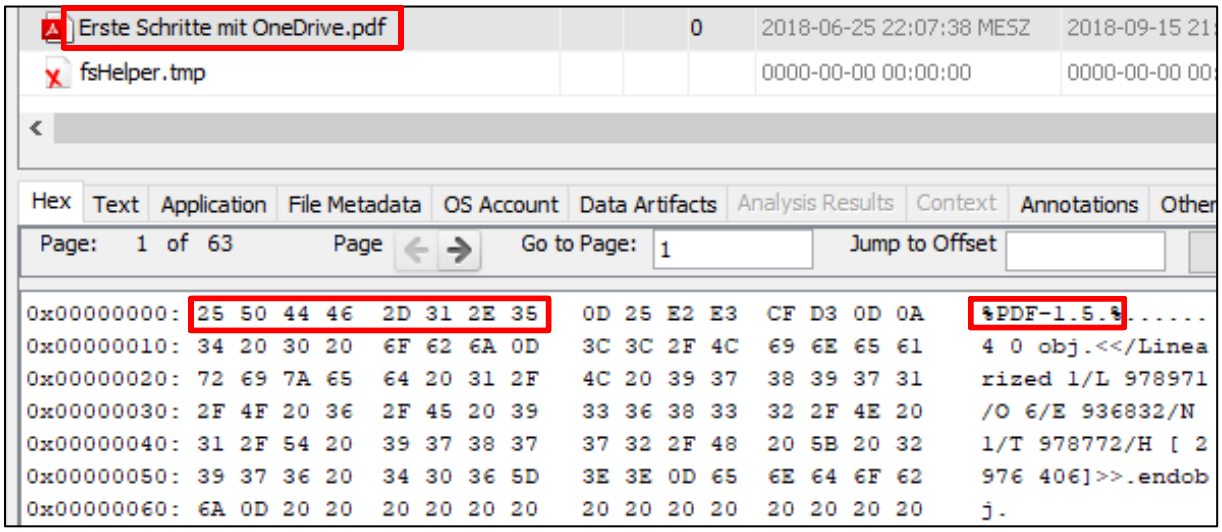


Abbildung 2.23 Header Signatur für das PDF-Format

Internet Media Typen

Eine besondere Gruppe bilden Dateitypen aus dem Internetbereich. Der Internet Media Type, auch MIME (Multipurpose Internet Mail Extension)-Type genannt, klassifiziert die Daten im Rumpf einer Nachricht im Internet. Bei einer HTTP-Übertragung wird z.B. einem Browser mitgeteilt, welche Daten der Webserver sendet – ob es beispielsweise ein Plain-Text-Dokument, ein HTML-Dokument oder ein PNG-Bild ist. Auch in E-Mails wird das „Content-Type“-Header-Feld dazu verwendet, die verschiedenen Daten zu klassifizieren [15]:

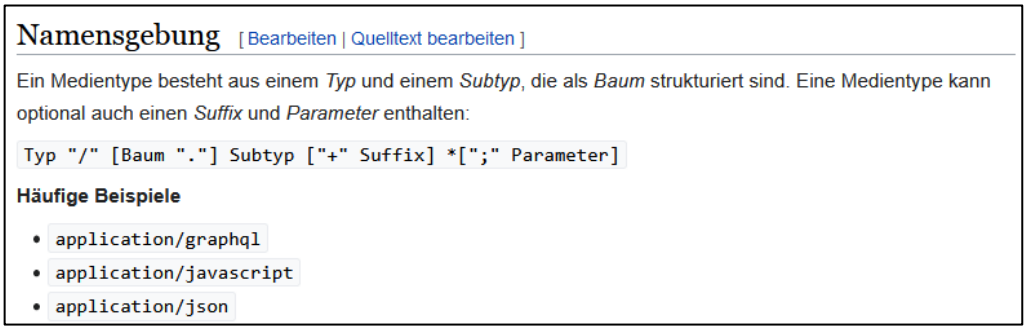


Abbildung 2.24: Aufbau eines MIME-Types

[15] Quelle: https://de.wikipedia.org/wiki/Internet_Media_Type

Erkennung und Extraktion der Meta- und Inhaltsdaten

Die nachfolgenden definierten Fähigkeiten beschreiben zwei Funktionen, wie verschiedene Formen von Text erkannt und extrahiert werden. EDiscovery-Anbieter verwenden für den Zugriff auf diese Dateiinhalte spezialisierte Programme wie z.B. dtSearch oder Tika. Leitfaden hierfür ist das EDRM-Dokument *3.0 Text-, Metadaten- und Bildextraktion* [16].

Verschlüsselte und beschädigte Dateien erkennen

Aus beschädigten oder verschlüsselten Dateien können weder Inhaltstext noch Metadaten extrahiert werden. EDiscovery-Systeme sollten diese „Nichtverarbeitung“ in einem Protokoll festhalten, damit deutlich wird, dass es neben dem Endergebnis auch noch weitere Daten gibt, die einen gesonderten Untersuchung zugeführt werden müssen.

Codierung erkennen

Dateien können unterschiedliche Zeichensätze (Codierungen) enthalten. Der vielleicht Bekannteste ist der ASCII-Zeichensatz, der 128 Zeichen aus dem englischen Alphabet (Groß- u. Kleinbuchstaben), Zahlen von 0-9 sowie Sonderzeichen unterstützt. Im universellen Codierungsstandard Unicode werden in der aktuellen Version 13 über 160 moderne und historische Sprachen unterstützt. Textextraktionssysteme von eDiscovery-Systemen müssen erkennen, welche Codierung in der Datei vorliegt, da sonst die folgende Extraktion und spätere Indexierung teilweise oder ganz falsch und damit unbrauchbar ist.

Sprache erkennen

Für die Inhaltsextraktion bei Dokumenten ist die Sprachidentifikation wichtig. Denn nur dadurch werden die richtigen Filter für Textextraktionen, Tokenisierung und diakritische Behandlung [17] eingesetzt. Zudem besteht dadurch die Möglichkeit der maschinellen Übersetzung in eine bekannte Sprache.

[16] Quelle: https://edrm.net/wiki/3-0text-metadata-and-image-extraction/

[17] Als *diakritische* Zeichen werden die Häkchen, Striche und Kringel über verschiedenen Buchstaben bezeichnet Quelle: https://de.wikipedia.org/wiki/Diakritisches_Zeichen

Metadatenextraktion

Die Verarbeitungsengine eines EDiscovery-Systems muss die grundlegenden Metadaten aus allen erkannten Dateien extrahieren können. Insbesondere interne Informationen über diese im Betriebssystem verwaltenden Dateien sollten Teil des Extraktionsprozesses sein. Bei E-Mail-Dateien sollte die Verarbeitungsssoftware nicht nur Metadaten aus der Datei selber, sondern auch Informationen aufnehmen, die aus der E-Mail-Datenbank (z.B. MS Exchange oder Office 365) gesammelt wurden.

Bildextraktionen

Die Verarbeitungs-Engine von eDiscovery-Systemen muss in der Lage sein, in Office-Dokumente wie auch in E-Mail-Nachrichten eingebettete Tabellen und Bilder (sog. embedded Objects) ähnlich wie E-Mail-Anhänge zum eigentlichen Textinhalt zu behandeln. Eine mögliche Abtrennung von derartigen Objekten zum Ursprungstext verändert das Original zum Endprodukt (Verarbeitungsergebnis) und mindert erheblich seine Beweiskraft. Zudem führt eine derartige Abtrennung (Bildextraktion) zu einer vermehrten Anzahl irrelevanter Datensätze, die manuell überprüft werden müssen.

Verarbeitung von SMS- und IM-Nachrichten aus Mobilgeräten

Short Messaging Service- und Instant Message-Nachrichten haben in der IT-forensischen Welt eine große Bedeutung. Sie enthalten erfahrungsgemäß häufig verfahrensrelevante Inhalte, sind aber in der Verarbeitung anders zu behandeln, da sie keine Dokumentdatentypen sind und sich somit Meta- und Inhaltsdaten nicht nach den beschriebenen Verfahren extrahieren lassen.

Derartige Daten liegen in proprietärer Form, häufig in verschlüsselten Datenbanken vor und unterliegen damit völlig anderen Gesetzmäßigkeiten. Proprietäre Software schränkt bekanntermaßen die Wieder- bzw. Weiterverwendbarkeit in Standard IT-Systemen (was eine Verarbeitungs-Engine voraussetzt) stark ein.

Aus Apps von Mobilgeräten lassen sich somit nur Daten verarbeiten, die dem Dateityp-Prinzip entsprechen, beispielsweise Anrufprotokolle, Kontakte, Kalenderelemente oder Voicemails.

Diese technische Gegebenheit ist der Grund dafür, dass eDiscovery-Systeme wie NUIX oder AXIOM nur in begrenztem Umfang oder gar nicht in der Lage sind ausge-

lesene Rohdaten aus Smartphones aufzubereiten. Zum Zeitpunkt der Fertigung dieser Master Thesis hat der Mobildevice Forensic Softwarehersteller CELLEBRITE ebenfalls ein eDiscovery-Produkt namens PATHFINDER auf den Markt gebracht, was sich darauf konzentriert genau diese Schwierigkeit zu lösen und damit dem Ziel, eine allumfassende eDiscovery-Lösung für alle digitalen Beweismittel anzubieten, näherkommt als seine Konkurrenten.

Daten aus Social Media- und Collaborations – Plattformen

Relevanz können auch Daten aus Social Media- oder Collaborations-Plattformen haben. Es handelt sich um Daten, die von verschiedenen interaktiven sozialen Plattformen generiert werden, die oft eine Messaging-/Chat-Komponente oder andere Kommunikation, wie z.B. Social-Media-Posts oder Filesharing enthalten. Anwendungsbeispiele hierfür sind im sozialen Bereich Facebook, Instagramm oder Twitter und für berufliche Zwecke MS Teams, Slack oder auch Google Chat, die nicht nur Kommunikation, sondern auch den Austausch oder die gemeinsame Nutzung von Dateien ermöglichen.

Dies geschieht durch die Sicherung ganzer Website-Inhalte. Entscheidend ist, dass die Erfassungs- bzw. Sicherungssoftware von Drittanbietern Datei- und Textformate verwendet, die vom eDiscovery-System erkannt und für die Weiterverarbeitung unterstützt werden. Denn nur so kann die notwendige Indexierung durchgeführt werden, was wiederum erfolgreiche Suchen ermöglicht.

Fazit

Auch in diesem Prozessschritt gibt es große Parallen in der Funktionalität zu IT-Forensik-Tools. Dennoch finden hier die ersten Unterschiede im Funktionsumfang statt, denn hier werden die Daten nicht nur extrahiert und in ein auswertbares Format gebracht, sondern für die Geeignetheit der Ausles- und Übertragbarkeit in das Verarbeitungsergebnis analysiert.

Bei der Suche nach einem geeigneten eDiscovery-System ist es wichtig zu wissen, welche der Container-Formate von der Verarbeitungs-Engine des eDiscovery-Systems unterstützt werden. Wird z.B. das selten gewordene aber noch immer vor allem in größeren Institutionen (z.B. Bundeswehrverwaltung, Stand 2022) wegen seiner hohen Sicherheitsstandards verwendete IBM-Mail-System Lotus Domino Server/Lotus Notes Client mit seinem Containerformat *.NSF von dem angeschafften eDiscovery-

System nicht unterstützt, so hat man sich für ein Produkt entschieden, dass unter Umständen die eigenen Anforderungsbeschreibungen im Hinblick auf die Auswertung von E-Maildateien nicht erfüllt. Dies gilt allerdings im gleichen Maße für klassische IT-Forensik Tools.

Eine weitere wichtige Fähigkeit von eDiscovery-Lösungen ist es, eine qualitativ hochwertige Texterkennungs- und -extraktions-Engine zu haben. Ganz besonders wichtig wird diese Fähigkeit bei Dateninhalten aus unterschiedlichsten Sprachkreisen. Sie entscheidet darüber, ob die Text-Extraktionen aus den verfahrensrelevanten Dateien korrekt und auch vollumfänglich stattfinden. Eine fehlerhafte oder unvollständige Textextraktion liefert eine unvollständige Indexierung, weil sie die Metadaten oder Textinhalte nicht erkennt, was wiederum Voraussetzung für ein erfolgreiches Finden und Suchen im Datenbestand ist.

2.6 Erzeugen

Im 3. Verarbeitungsschritt geht es um die Weiterverarbeitung extrahierter Daten. Hier werden die erkannten und extrahierten Meta- und Inhaltsdaten in eine oder mehrere Datenbanken überführt und damit für die späteren Recherchen und Suchen verfügbar gemacht. Diesen Prozess kann man auch als „Harmonisierung der Datenformate“ bezeichnen.

2.6.1 Extraktion und Normalisierung von Text

In einem ersten Schritt der Textextraktion wird von der Verarbeitungs-Engine entschieden, wie der Text nachfolgend indexiert [18] wird. Dies erfolgt durch eine „Normalisierung“ der Textinhalte. Die wichtigsten Methoden hierzu sind die:

- **Normalisierung der Groß- und Kleinschreibung** durch Konvertierung nur noch in Kleinschreibung

- **Didaktische Normalisierung**, indem dafür gesorgt wird, dass bei Sprache bedingter Sonderzeichen wie 'æ' (Beispiel aus dem Dänischen) durch Um-

[18] Quelle wikipedia: Indexierung = Zuordnung von Deskriptoren zu einem Dokument zur Erschließung der darin enthalten Sachverhalte

wandlung in ae die Trefferquote bei Recherchen bzw. Suchen hoch bleibt

- **Unicode-Normalisierung**, die z.B. dafür sorgt, dass sprachspezifische Akzentzeichen wie é in ein ASCII-Äquivalent wie *e* und den Akzent *aigu* aufgetrennt werden und damit recherchefähig bleiben

- **Normalisierung der Zeitzone**, durch Angleichung unterschiedlicher Zeitzonen-Informationen in einen einzigen Standard, da es ansonsten gerade bei E-Mail-Nachrichten aus unterschiedlichen Zeitzonen zu Irritationen bzw. falschen Interpretationen kommen würde, wenn Messages z.B. aus anderen Zeitzonen wie Pacific Standard Time zugesandt werden, und man den ursprünglichen Timestamp (UTC -11 = Hawaii) ohne Relation zur Zeitzone der gesicherten Datei (UTC +1 = Deutschland) darstellen würde

- **Text-Tokenisierung**, bei der ein Token, auch gen. lexikalische Einheit, im Suchindex platziert wird. Dies geschieht durch Trennung der Wörter, indem die meisten Satzzeichen entfernt und die Leerzeichen zwischen den einzelnen Token verwendet werden, um sie als separate Einheiten zu definieren. Bei bestimmten Sprachen wie Japanisch gibt es keine Leer- und Trennzeichen, hierfür werden spezielle Tokenisierungsprogramme verwendet, um damit eine erfolgreiche Indexierung zu ermöglichen.

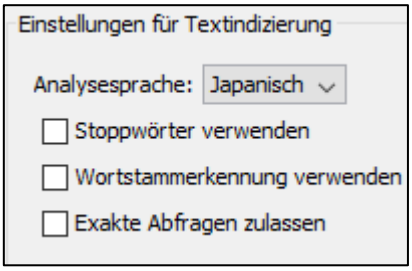


Abbildung 2.25: Text-Tokenisierung unter NUIX WS 8.8

- **Optionale Einstellungsmöglichkeiten der Wortnormalisierung**, die es dem Anwender ermöglichen zu wählen, ob z.B. Standardsatzzeichen wie “” oder & ebenso indexiert werden sollen oder nicht. Werden diese Satzzeichen mit indexiert, erhöht sich die Verarbeitungszeit und das Speichervolumen für die Indexierung. Andererseits kommt es zu keinen Suchtreffern bei Abfragen wie AT & T, wenn die Einstellung deaktiviert ist. Hier sollte ein eDiscovery-System dem Nutzer die Wahl fallabhängig überlassen.

- Office-Dateien können Kommentare enthalten, von Lesern, die z.B. Anmerkungen an ein Protokoll zu einer Besprechung anfügen. EDiscovery-Systeme sollten in der Lage sein, auch diese zusätzlichen Informationen neben dem Ur-

sprungstext zu extrahieren.

2.6.2 Ausgabeformate

Bei dieser Thematik geht es um die Ausgabeformate, in denen die verfahrensrelevante ESI an den Prozessgegner übergeben werden müssen. Im EDRM-Dokument *Produktionsleitfaden* [19] wird in der Bundeszivilprozessordnung Regel 26f darauf verwiesen:

„...dass die Methode und das Format, nach denen ESI erstellt werden sollen, von den Parteien frühzeitig im Ermittlungsprozess geprüft und ausgehandelt werden sollten..."

In deutschen Strafverfahren spielt diese Regelung und auch für die Aufgabenstellung dieser Master-Thesis keine Rolle. Es ist im deutschen Strafprozessrecht nicht festgelegt, dass sich die Prozessparteien auf gemeinsame Ausgabeformate einigen müssen. Daher wird nur kurz auf die wichtigsten Begriffe bei dieser Grundlagenvermittlung eingegangen, um das Verständnis zu steigern, denn die Begriffe finden immer wieder Anwendung in den Bedienungsanleitungen von eDiscovery-Programmen:

- Von **nativen Dateiformaten** beim Austausch der Daten wird gesprochen, wenn die Ursprungsformate wie beispielsweise Word, Excel oder Adobe Acrobat erhalten bleiben. E-Mail-Dateien entsprechen in gewisser Weise Datenbanken, durch ihre vielen Unterverzeichnisse. Ein Beispiel für eine native Produktion kann daher vorliegen, wenn aus der PST-Datei Einzelnachrichten in das native *.MSG Format konvertiert werden. Auch wenn Tabellen über die üblichen Druckformatgrenzen wie A4 oder A5 hinausgehen oder die enthaltenen Formeln eine Relevanz beinhalten, wählt man native Dateifomate in der Übergabe.

- Bevor z.B. E-Mail-Daten in ihrer Gesamheit übergeben werden, wird geprüft, ob die Inhalte auch nicht relevante Teile enthalten, die abgetrennt werden müssen. Für diese Dokumentenprüfung findet dann die Extraktion und Konvertierung in Einzeldateien statt. Dies sind z.B. strukturierte Textformate wie *.html oder *.xml, die sich so einfacher indexieren lassen. Diese neuen Einzeldaten gehören zur Gruppe der **near (nahezu) native** Dateiformate. Besonders bei Datenbanken mit Verfahrensrelevanz wird dieses Dateiformat genutzt, da zumeist nicht die gesamte Datenbank, sondern nur bestimmt Auszüge, Datenfelder oder Tabellen benötigt

[19] Quelle: https://edrm.net/resources/frameworks-and-standards/edrm-model/production/

werden, die dazu wiederum in die bereits o.g. Einzeldateien konvertiert werden.

- Electronic Discovery Anwendungen besitzen auch die Möglichkeit ESI in Bild- oder Papierformaten zu erstellen. Man nennt dieses Format **near paper (Bildnah).** Dies geschieht z.B. durch das “Rendern” oder Einscannen von Papierunterlagen in Bildformate.

Fazit

In dem letzten Verarbeitungsschritt des Processings wird der große Unterschied zu klassischen Forensik-Tools deutlich. Hier wird ein spezifischer Konvertierungsprozess zur Übergabe in ein neues Produktionsergebnis eingesetzt. Dabei ist der Umwandlungsprozess darauf ausgerichtet, alle textbasierenden Informationen dieser analysierten und erkannten Datenformate auszulesen und zu indexieren. Zusätzlich kann bei eingescannten papiernen Dokumenten über optische Zeichenerkennungstechniken (OCR) der enthaltene Text indexierungsfähig gemacht werden. Auch wenn diese Indexierungsfunktion bei klassischen IT-Forensik-Tools vorhanden ist, so ist deren Leistungsfähigkeit bei eDiscovery Programmen erheblich besser, denn genau für die Verarbeitung und Verwaltung großer Datenmengen sind derartige Anwendungen programmiert worden.

In der Praxis wird dieser Unterschied deutlich, wenn man versucht eine größere Anzahl von Images unter einem Fall (Case) in den bewährten Forensik-Tools Encase oder X-Ways zu bündeln, um über diese Gesamtdatenbestände Suchen nach bestimmten Begriffen oder Worten durchzuführen. Auch bei entsprechend leistungsfähiger Hardware ist häufig ein „Absturz“ oder eine sehr lange Funktionsbelegung (Sanduhr läuft) dieser Tools zu beobachten.

Bei eDiscovery-Anwendungen tritt dieses Phänomen des Programmabsturzes nicht auf. Zwar wird mit zunehmender Datenmenge irgendwann auch eine eDiscovery-Anwendung langsamer, doch nicht in dem Ausmaß eines typischen IT-Forensik-Tools. Zudem kann mit einer Zusatzanwendung wie der Elasticsearch-Backend-Datenbank gerade bei großen zu verarbeiteten Datenbeständen die Leistungsfähigkeit bei Suchen nochmal deutlich gesteigert werden.

2.7 Export

Wie bei forensischen Tools auch, müssen eDiscovery-Systeme eine Export- bzw. eine Ausgabefunktion besitzen, damit die Möglichkeit besteht, die Inhalte einer Word-Datei an Dritte weiterzugeben und mit anderen Tools weiterzuverarbeiten. Durch den Selektionsprozess, was Relevanz besitzt oder nicht, stehen nur die Daten aus dem Verarbeitungsprozess zur Verfügung und nicht mehr das Original-Format.
Für die Realisierung werden daher sog. Ladedateien (Load files) verwendet. Eine Ladedatei organisiert einen Dokumentenkorpus und die dazugehörigen Metadaten und schreibt über diese Funktion die im eDiscovery-System gesammelten Daten in eine Datenbank zur Übergabe an die gegnerische Partei. Im EDRM Dokument *4.0 Verarbeitungsleistung* [20] unter *4.2 Entwickeln von Ladedateien* wird hierzu folgendes bemerkt:

„...Die Datenbank muss geeignete Informationen enthalten, um jeden Datensatz mit den zugrunde liegenden nativen, Bild- und Textdokumenten zu verknüpfen, die während der Verknüpfung ausgegeben werden...Daher muss die Ausgabe eines Verarbeitungssystems eine Ladedatei enthalten, die wichtige Informationen zu jedem Dokument liefert und auch das Laden von Daten und Dateien in das System erleichtert.“

Typische Standardproduktionsexportformate sind DAT, CSV, LFP und OPT. Je nach verwendeter Software sind auch weitere Exportformate möglich, wie das Beispiel von NUIX Discover in der nächsten Abbildung 2.26 zeigt:

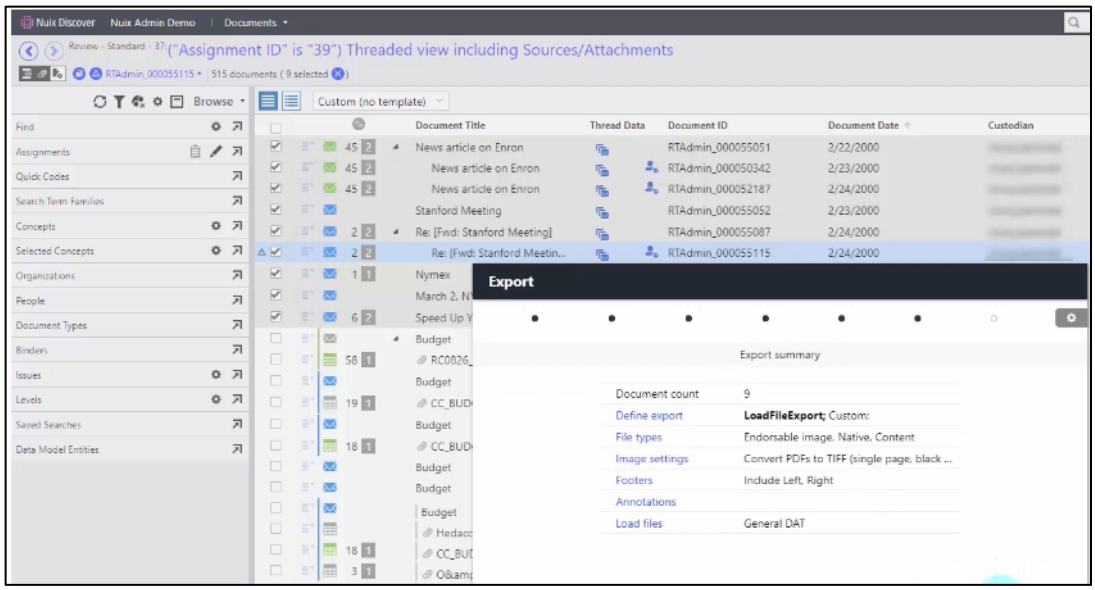
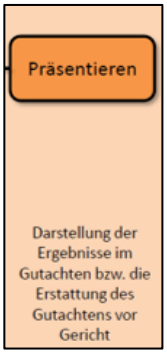


Abbildung 2.26: Exportmöglichkeiten unter NUIX

[20] https://edrm.net/wiki/4-0-processing-output/

Die Beachtung dieser Exportformate in deutschen Strafverfahren spielte bisher keine Rolle, sie ist nur zwingend in den US-Federal Rules of Civil Procedure. Daher wird auch nicht näher auf diese Thematik eingegangen.

2.8 Präsentation



Dieser letzte Prozessschritt im EDR-Modell geht weniger auf technische, sondern auf formale und juristische Bedingungen sowie Definitionen im US Prozessrecht ein. Hier wird z.B. definiert was Experten, erforderliche Ausstattungen von Gerichten, Anhörungen und Anträge, Vorbereitung von Exponaten sind.
Dieser Bereich wird ebenfalls nicht näher ausgeführt, denn er hat für die Aufgabenstellung dieser Master-Thesis (technische Eignung von eDiscovery-Anwendungen in deutschen Strafverfahren) keine Relevanz.

2.9 Berichterstattung

Die Verarbeitung von Originaldaten und Konvertierung in ein einheitliches Endprodukt durch die eDiscovery-Anwendung muss jederzeit nachvollziehbar sein. Dies erfolgt durch die automatische Erstellung von Protokollen im Verarbeitungsprozess. Der EDRM-Leitfaden *5.0 Reporting* [21] bemerkt hierzu folgendes:

2.9.1 Dateiinventar-Berichte

EDiscovery-Systeme sollten Protokolle erstellen, aus denen

- die Anzahl der auf einem bestimmten Datenträger enthaltenen Dateien
- die Art der auf dem Datenträger enthaltenen Dateien und
- die Größe der auf dem Datenträger enthaltenen Daten

hervorgeht. Zudem sollten Verzeichnislisten der Dateinamen angelegt werden.

2.9.2 Verwahrstellen-Berichte

Verwahrstellen-Berichte sollten

- den Namen des Verwalters
- empfangene und verarbeitete Datensätze
- Dateidaten, Typen und Größen sowie

[21] https://edrm.net/wiki/5-0-reporting/

- Ausnahmeinformationen für Dateien, die nicht verarbeitet werden konnten, enthalten.

2.9.3 Filter-Berichte

Hier sollte dokumentiert werden, welche Dateien oder auch erkannten Viren durch eine Filterung entfernt wurden.

2.9.4 Bericht zur Dateiverarbeitung

Eine eDiscovery-Software sollte alle Dateiverarbeitungsschritte aufzeichnen, die der Chain of Custody Regel dienlich sind.

2.9.5 Ausnahmeberichterstattung

Dateien, die nicht verarbeitet werden können, sollten im Endprodukt, also der aufnehmenden Datenbank, als Ausnahmen gekennzeichnet sein. Dies sind Dateien, für die

- kein Text oder Metadaten extrahiert oder
- keine Abbildungen gerendert

werden konnten. Im Idealfall sollte im Protokoll der Grund (Failure), warum die Datei nicht verarbeitet wurde, enthalten sein.

2.10 Mehrwerte – Big Data-Analysen

Die Fähigkeit, große Datenmengen aufzunehmen, zu verarbeiten und unterschiedlichste Datenformate zu harmonisieren, bietet eine ideale Grundlage für Big Data-Analysen. Zwar findet sich im EDR-Modell hierzu kein Regelwerk, welches diese Funktionalität definiert, doch in allen großen bekannten eDiscovery-Tools ist dieser Mehrwert enthalten. Dazu gehören auch die eDiscovery-Tools NUIX und in kleinerem Umfang AXIOM, welche später Gegenstand verschiedener Tests sind.

Big Data-Analysen bezeichnen einen Sammelbegriff für Softwareentwicklungen, die in der Lage sind, „...*Datenmengen, welche zu groß und zu schnellebig oder zu schwach strukturiert sind, um sie mit manuellen herkömmlichen Methoden der Datenverarbeitung auszuwerten...*“ [22]

[22] https://de.wikipedia.org/wiki/Big_Data

Im nachfolgenden Beispiel von NUIX werden gleichartige Datenbestände (hier E-Mail-Daten) aus unterschiedlichen Sicherungen und damit Nutzern verglichen und zeigen die Kommunikationsverbindungen zwischen verschiedenen Nutzern durch eine Visualisierung an:

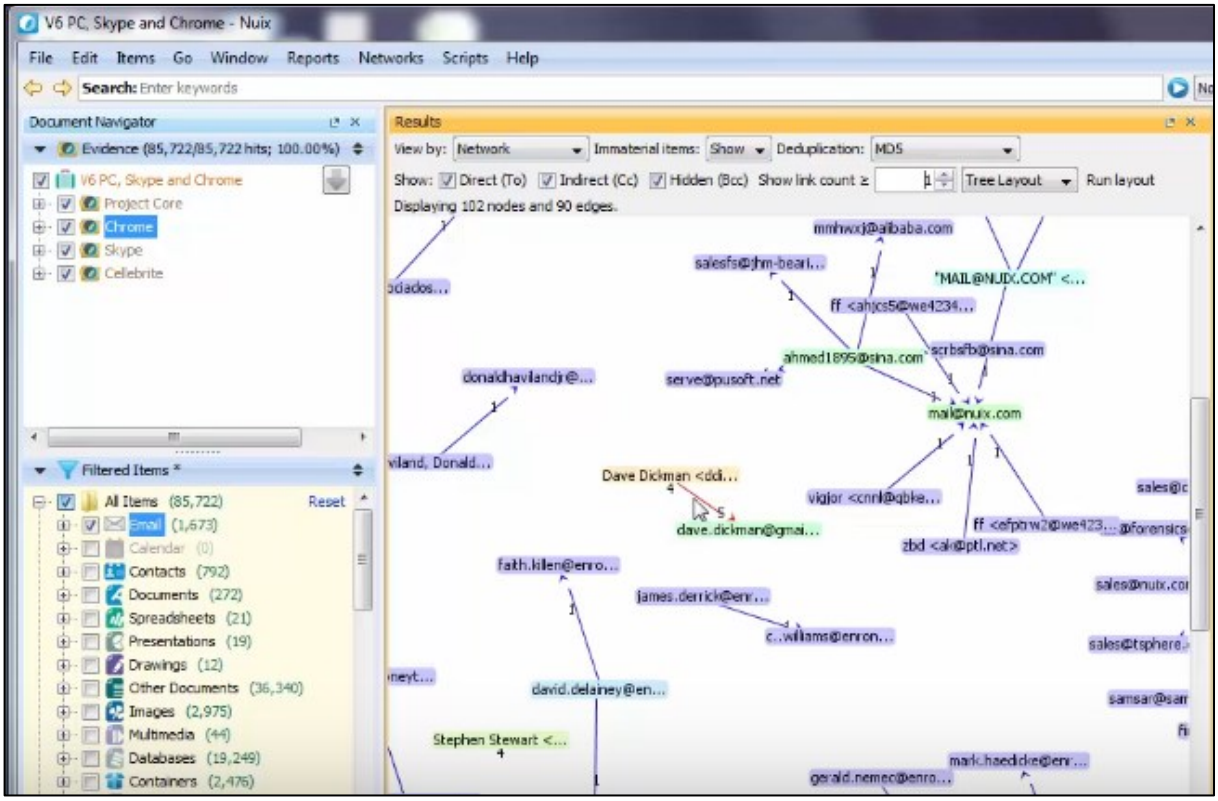


Abbildung 2.27: Darstellung von Kommunikationsbeziehungen unter NUIX

Ebenso können Daten aus Mobilgeräte-Sicherungen nach Geokoordinaten suchen und zusammen mit einer Zeitstempelanalyse ein Bewegungsbild von einem oder mehreren Smartphone-Nutzern erstellen und auf einer Karte visualisieren:

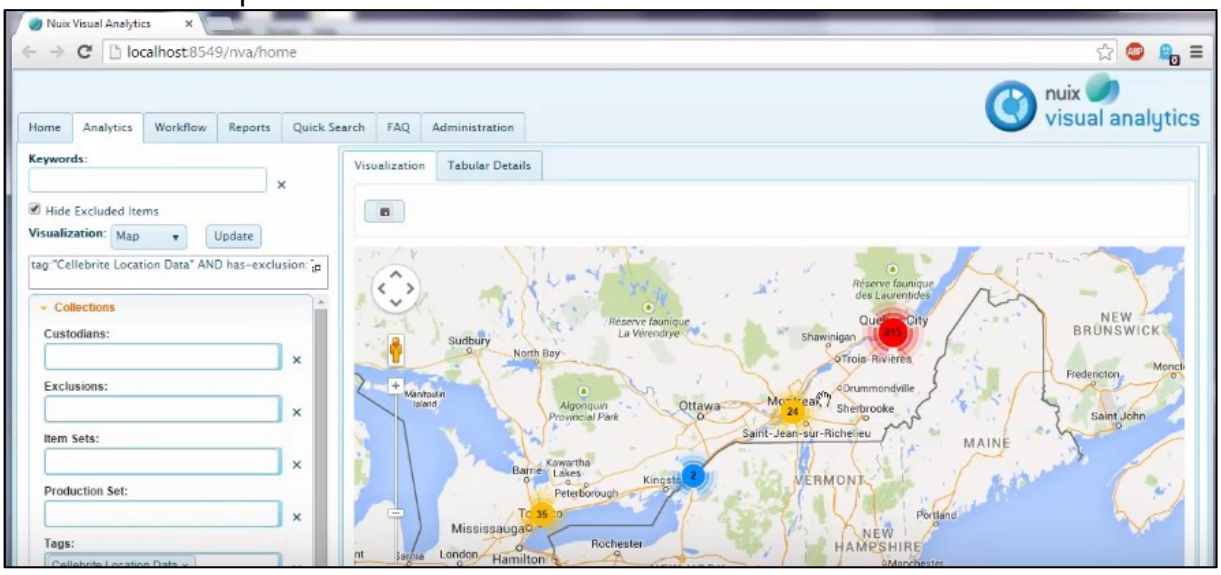


Abbildung 2.28: Bewegungsbild anhand von Geokoordinaten aus einem Smartphone

2.11 Zusammenfassung

Das Ziel eines eDiscovery-Systems ist es, Nutzer-Daten aus den unterschiedlichsten Anwendungsbereichen wie z.B. Dokumenten, E-Mails, Instant Messengern so aufzubereiten, dass in dem Endprodukt über den Gesamtbestand Auswertungen, Recherchen und Analysen möglich werden. Dies macht die Auswertung ggü. den klassischen Arbeitsweisen enorm effizient, da nicht für jede Datenaufbereitung die Recherche nach bestimmten Begriffen, Namen oder Zahlenreihen wiederholt werden muss.

Dabei muss die Verarbeitungs-Engine in dem Konvertierungsprozess gleichzeitig dafür sorgen, dass die inhaltliche Authentizität der Originaldaten weder verändert noch in dem späteren Verarbeitungsergebnis „vergessen“ also gar nicht angezeigt wird.

Das Grundprinzip einer eDiscovery-Anwendung im Vergleich zu klassischen IT-Forensik-Tools besteht darin, dass aus erkannten Datentypen die Meta- und Textinhaltsdaten mit seinen verschiedenen Textformaten erkannt, ausgelesen und in entsprechende Spalten einer oder mehrerer Datenbanken kopiert werden. Das ist aufgrund unterschiedlicher Textformate und einer großen Anzahl unterschiedlicher Sprachen und deren zusätzlichen Sprachzeichen (z.B. Apostroph) besonders anspruchsvoll.

Als bedeutender Mehrwert und neue Technologie in der IT-Forensik sind die noch am Anfang stehenden Big Data-Analysefähigkeiten von eDiscovery-Anwendungen zu sehen. Bereits jetzt ist zu erkennen, dass diese Technologie das Maß der Aufklärungsmöglichkeiten bei Straftaten einen erkennbaren Schritt nach vorne bringt.

Neben diesen positiven Effekten gibt es aber auch einen „Pferdefuß“. Electronic Discovery-Systeme können aufgrund ihrer Verarbeitungsmethode von textbasierenden Inhalten nur Lösungen für solche Delikte bzw. Fälle bieten, bei denen die verfahrensrelevante Spur aus einer Text basierenden Information besteht. Nur weil alle Suchbegriffe nicht zum Erfolg geführt haben, heißt das nicht, dass die Datensicherungen bzw. -aufbereitungen keine Relevanz haben. Erst durch das Verstehen der installierten Programme oder Datenverzeichnisse zu einem gesicherten Gerät können Hinweise und Ideen entstehen, wie ein Täter vorgegangen ist und bieten Ansätze für weitere Untersuchungen mit anderen Suchbegriffen.

3 Untersuchungen

Nachdem im Abschnitt 2 die grundlegende Funktionsweise eines eDiscovery-Systems erklärt wurde, wird nun der Fragestellung nachgegangen, ob die Methoden von eDiscovery-Systemen Bestand vor Strafgerichten haben. Dies erfolgt durch die nähere Betrachtung und den Vergleich mit den klassischen IT Forensik-Methoden und Arbeitsabläufen. Dabei werden sowohl die formalen als auch technischen Unterschiede betrachtet.

Bei den technischen Unterschieden wird die Fragestellung anhand von ausgewählten Testszenarien überprüft, ob eDiscovery-Anwendungen aufgrund ihrer Methodik der Datenverarbeitung und -aufbereitung Defizite bzw. Risiken im Bezug auf die Qualität der Anforderungen im Beweismittelrecht für deutsche Strafverfahren beinhalten.

3.1 Formaler Vergleich zwischen dem EDRM- und SAP-Modell

Für einen formalen Vergleich der klassischen IT-Forensik-Untersuchungsarbeit und eDiscovery-Systeme für IT-forensische Aufgaben bietet sich die Gegenüberstellung der definierten Anforderungen, der jeweiligen Ziele dieser Modelle, Vorgehensweisen und den dafür notwendigen Prozessschritten und deren Funktionsumfänge an.

3.1.1 Anforderungen

Um eDiscovery-Systeme offiziell auch für deutsche Strafverfahren einsetzen zu können, müssen sie genauso wie die klassischen IT-Forensik-Tools deren Anforderungen erfüllen. Das Bundesamt für Sicherheit in der Informationstechnik hat in seinem *Leitfaden IT Forensik* [23] folgende Anforderungen an die Vorgehensweise definiert:

Akzeptanz

Die von Ermittlern angewandten Methoden und Schritte müssen in der Fachwelt beschrieben und allgemein akzeptiert sein. Der Einsatz neuer Verfahren und Methoden ist zwar prinzipiell nicht ausgeschlossen, jedoch sollte ein Nachweis der Korrektheit erfolgen.

[23] https://www.bsi.bund.de/DE/Themen/Oeffentliche-Verwaltung/Sicherheitspruefungen/IT-Forensik/forensik_node.html

Bewertung: EDiscovery-Systeme werden seit mehr als 15 Jahren in den USA angewendet und haben sich etabliert, was sich auch in den "*Guidelines for the Discovery of Electronically Stored Information"* des United States District Court, Northern District of California ausdrückt [24] und damit seine Akzeptanz in den US Behörden belegt. Deutsche Wirtschaftsprüfungsgesellschaften wie KPMG oder Deloitte setzen eDiscovery-Anwendungen bereits seit mehr als 10 Jahren in Deutschland ein. Seit ca. 8 Jahren nutzen zunehmend deutsche Sicherheits- und Kontrollbehörden des Bundes und der Länder ebenfalls eDiscovery-Systeme. Mehrere bedeutende Strafverfahren (Steuerstrafverfahren gegen den ehemaligen FCB-Präsidenten Hoeneß, Dieselgate Daimler-Benz und aktuell Dieselgate VW) wurden und werden mit derartigen eDiscovery-Systemen bearbeitet und für die Anklage vorbereitet.
IT-Forensik-Tools wie Encase oder X-Ways wurden in bestimmten Versionsständen vom FBI und BKA evaluiert. Eine Evaluierung für weit verbreitete eDiscovery-Tools wie z.B. NUIX ist dem Autor bisher nicht bekannt geworden und auch nicht über Google-Recherchen zu finden. Andererseits sind bisher auch keine Fehlerhaftigkeiten bei Einsatz von eDiscovery-Systemen in Wirtschaftsprüfungsgesellschaften oder deutschen Strafprozessen nachgewiesen bzw. bekannt geworden.
Das in diesem Bereich führende eDiscovery-Tool NUIX erfüllt seit 2019 die ISO Norm 27001 [25]. Sie bezieht sich in erster Linie auf das Informationssicherheitsmanagement einer Softwareanwendung und weniger auf den Nachweis einer korrekten Verarbeitung, belegt aber, dass es dem Softwareanbieter wichtig ist, nachzuweisen, dass das Unternehmen vor *unbefugtem Zugriff oder Kompromittierung geschützt* und das IT-Personal sich als *gewissenhaft und sachkundig in Best Practices erwiesen hat.* Zusammenfassend wird das Merkmal Akzeptanz als erfüllt angesehen.

Glaubwürdigkeit

Hier geht es um die Nachweisbarkeit der Funktionalität und Robustheit der Methoden in einer Anwendung. Diese sollten bei Bedarf dem Gericht nachgewiesen werden können. Hierzu bemerkt der anerkannte IT-Forensik Experte und Autor Alexander Geschonneck in seinem Buch "Computer Forensik" [26] folgendes:

[24] https://cand.uscourts.gov/filelibrary/1117/ESI_Guidelines-12-1-2015.pdf
[25] https://www.nuix.com/news/nuix-achieves-isoiec-270012013-certification
[26] ISBN 978-3-89864-534-8, 3. Auflage, S. 63

“Es ist sicherlich immer schwierig, wenn man irgendein Tool mit Daten füttert und dann am Ende irgendwelche Ergebnisse herauspurzeln, deren Zustandekommen nicht nachvollziehbar ist…”

Bewertung: Auch klassische IT-Forensik-Tools haben eine Teilautomatisierung, wenn es darum geht, Datenextraktionen oder Dateianalysen für die Bestimmung des Erstellers, Nutzers oder der Zeitstempel vorzunehmen. Doch hierbei findet keine Veränderung statt. Die Rohdaten bleiben unverändert und können über einen HEX-Editor eingesehen werden. Damit besteht eine Nachvollziehbarkeit, wenn z.B. aus den Rohdaten einer Datei der Zeitstempel errechnet wird, was normalerweise vom Forensik-Tool geliefert wird.
EDiscovery-Systeme entnehmen Daten nach bestimmten Mustern, schreiben diese in eine Datenbank, verändern diese dabei, die dann als Basis für die Auswertung durch den Ermittler dient.
Um wirklich beurteilen zu können, ob ein eDiscovery-System korrekt arbeitet, müßte sein Quellcode offengelegt werden oder man vergleicht die Verarbeitungsergebnisse einer eDiscovery-Anwendung mit unveränderten Originaldaten über ein klassisches IT-Forensik-Tool (z.B. ob der richtige Zeitstempel berechnet wurde oder alle Originaldaten vollständig verarbeit und im Verarbeitungsergebnis fehlerfrei dargestellt werden). Schafft es die Verteidigung, Zweifel durch den Nachweis einer Fehlerhaftigkeit in dem eingesetzten eDiscovery-Tool zu begründen, ist die Glaubwürdigkeit aufgehoben. Ob dieses Merkmal erfüllt wird, zeigen die nachfolgenden Tests unter 3.2.

Wiederholbarkeit

Die im gesamten Ermittlungsprozess verwendeten Methoden und Hilfsmittel müssen bei Anwendung von Dritten wiederholbar sein. Das bedeutet, dass eine dritte Person, die die gleichen Schritte durchführt, zum gleichen Ergebnis kommen muss.

Bewertung: Da bei dem Nachweis der Wiederholbarkeit der Untersuchungsergebnisse die gleichen Originaldaten und die gleiche Softwareversion eingesetzt werden, und somit die gleiche Programmierung auf die Analyse und Verarbeitung der Daten angewendet wird, ist der Logik folgend kein anderes Ergebnis möglich. Dies ändert sich auch nicht durch eine andere eingesetzte Hardware oder andere Betriebssystemversionen für die Untersuchungsumgebung. Entscheidend für die Gleichheit der Ergebnisse sind die gleichen Programmversionen von eDiscovery-Systemen, gleichen

Konfigurationseinstellungen zum Processing und gleichen Culling-Anwendungen. Damit erfüllen eDiscovery-Systeme genauso wie IT-Forensik-Tools immer dieses Merkmal der Wiederholbarkeit, sofern gleiche Software-Versionen und Konfigurationen eingesetzt werden.

Integrität

Im Rahmen der Untersuchungen darf an den gesicherten Daten keine Veränderung vorgenommen werden. Es muss jederzeit die Integrität der Beweise nachweisbar sein. Bei klassischen IT-Forensikprogrammen wird beim Erstellen von forensischen Datensicherungen (Acquiring) jeweils über den Gesamtbestand des gesicherten Datenbestandes ein Hashwert (üblicherweise MD5 und/oder SHA1) gebildet, der beim Einlesen in das Forensik-Tool durch das sog. Verifying überprüft wird (siehe hierzu auch S. 12, Abbildung 2.7).

Bewertung: Alle professionellen IT-Forensik Tools besitzen eine Verifying - (Überprüfung des gebildeten Hashwertes) Funktion. Weder in den Bedienungsanleitungen zu den eingesetzten eDiscovery-Tools AXIOM und NUIX Workstation noch in den Protokolleinträgen zur Verarbeitung derartiger Images fanden sich Hinweise auf diese Überprüfungsfunktion (siehe hierzu Ergebnisse Test 1, unter 3.2.2, S. 58 - 59). Anfragen über die Supporthotlines dieser Softwareanbieter wurden bis zur Fertigung dieser Master Thesis nicht beantwortet. Auch eine “Google” – Recherche brachte hierzu keine Erkenntnis, ob eDiscovery-Systeme diese Prüffunktion ebenfalls unterstützen. Ggf. muss vor dem Einlesen dieser forensischen Images mit anderen Tools die Prüfsumme für das jeweilige forensische Image überprüft werden.

Ursache und Auswirkungen

Die für die Untersuchungen gewählten Methoden müssen es ermöglichen, logisch nachvollziehbare Verbindungen zwischen Personen, Ereignissen und Beweisspuren herzustellen.

Bewertung: Im Wesentlichen sind dies die Faktoren zur korrekten Zeitstempelberechnung, der Nachweis wer, Ersteller und Nutzer dieser relevanten Daten ist, und wo diese relevanten Spuren innerhalb des gesicherten Datenbestandes gefunden wurden. In den nachfolgenden Tests mit den zur Verfügung stehenden eDiscovery-Anwendungen

AXIOM und NUIX fanden sich diese Funktionen ebenso wie in den klassischen IT-Forensik-Tools.

Dokumentation

Jeder Schritt der Untersuchung muss angemessen dokumentiert werden können. Das heißt, alle eingesetzten IT-Forensik-Anwendungen für die Untersuchungen sowie wichtige Prozessschritte oder auch Zwischenergebnisse innerhalb dieser Programme müssen im Untersuchungsbericht erwähnt werden.

Bewertung: Klassische IT-Forensik-Tools wie X-Ways oder Encase legen bei jeder neuen Untersuchung ein Fallverzeichnis an, in dem Protokolle mit unterschiedlichem Umfang zu durchgeführten Analysen aber auch der Export von Dateien gespeichert werden (siehe hierzu auch S. 50 ff). Bei der Fertigung von Gutachten ist es neben dem Hinweis eingesetzter Tools außerdem üblich, die eingesetzten Methoden (z.B. Einsatz bestimmter Skripts) und Maßnahmen näher zu beschreiben und die Zwischenergebnisse abzuspeichern.

Die hier getesten eDiscovery-Tools besitzen umfängliche Protokollfunktionen, insbesondere NUIX. Beide Tools dokumentieren in ihren Protokollen jeden Verarbeitungsschritt und auch -fehler sowie getätigte Konfigurationen zum Processing.

3.1.2 Ziel- und Aufgabendefinition

SAP-Modell

Das SAP-Modell steht für eine Abfolge von Prozessschritten, die in der klassischen IT-Forensik angewendet werden. Die Abkürzung steht für **S**ecure-**A**nalyze-**P**resent. Die Ziele des SAP-Modells ergeben sich aus der Definition zur IT-Forensik. Eine einheitliche Begriffsdefinition hat sich hierfür noch nicht durchgesetzt. Am treffendsten erscheint dem Autor die “alternative” Beschreibung aus dem IT-Forensik Wiki der Hochschule Wismar [27], da diese im Gegensatz zu anderen Definitionen genereller gehalten wird und nicht nur auf IT-Vorfälle im Zusammenhang mit Computerstraftaten im engeren Sinne (sprich Hacking-Verfahren) abzielt:

“Die IT-Forensik ist ein Teilgebiet der Forensik. Die IT-Forensik behandelt die Untersuchung verdächtiger Vorfälle im Zusammenhang mit der Informationstechnik

[27] https://it-forensik.fiw.hs-wismar.de/index.php/IT-Forensik

und der Feststellung des Tatbestandes und der Täter durch Erfassung, Analyse und Auswertung digitaler Spuren.“

EDR-Modell

In der Einführungsbeschreibung der EDRM-Webseitenpräsenz [28] wird zur Aufgabenbeschreibung unter anderem folgendes bemerkt:

“*…Das Ziel eines eDiscovery-Einsatzes ist es, relevante Informationen für Gerichtsverfahren, Schiedsverfahren oder eine Anhörung sichtbar zu machen. Der Prozess beginnt mit der Identifizierung elektronisch gespeicherter Informationen (ESI), die für das Verfahren relevant sein können und endet mit der Erstellung von sofortiger auswertbarer, nicht privilegierter ESI* [29] *für eine anfragende Partei…”*

3.1.3 Vorgehensweisen

SAP-Modell

Das Bundesamt für Sicherheit in der Informationstechnik beschreibt in seinem aktuellen “Leitfaden IT Forensik” Version 1.0.1 die dafür notwendigen Prozessschritte:

Dieser Leitfaden unterteilt die Vorgehensweise einer forensischen Untersuchung in die folgenden sechs Abschnitte:

- strategische Vorbereitung;
- operationale Vorbereitung;
- Datensammlung;
- Untersuchung;
- Datenanalyse;
- Dokumentation

Abbildung 3.1: BSI Leitfaden IT Forensik - Prozessschritte

EDR-Modell

In der FAST DECTECT Grafik (siehe Abbildung 2.2 auf S. 8) wurden die Prozessschritte bereits dargestellt:

- Informationsmanagement
- Identifikation
- Aufbewahrung und Erfassen
- Verarbeiten, Überprüfen, Analysieren
- Erzeugen und Präsentieren

[28] https://edrm.net/wiki/introduction/

[29] Als privilegiertes ESI Material werden digitale Informationen bezeichnet, die von einer Prozesspartei auf Antrag nicht offen-gelegt werden, wenn sie als Prozessvorbereitungsmaterial deklariert werden. Alles andere Material wird als non-privileged (nicht privilegiertes) Material definiert und muss offengelegt werden. Quelle: Federal Rule Civil Prozess, Artikel 26 b, Artikel 5 Privile-gienansprüche

3.1.4 Prozessschritte und Funktionsumfänge

Informationsmanagement, Identifizieren, Aufbewahren und Erfassen

Die strategischen, operativen Vorbereitungshandlungen und Datensammlungen aus dem SAP-Modell sind mit dem Informationsmanagement und der forensischen Sicherung bzw. Akquise aus dem EDR-Modell vergleichbar und damit in der Vorgehensweise und im Funktionsumfang gleich.

Verarbeiten, Überprüfen und Analysieren

Beide Modelle müssen vor den eigentlichen Analysen Datenextraktionen an den Sicherungen vornehmen. Bei IT Forensik-Tools finden Extraktionen ebenso wie bei eDiscovery-Anwendungen immer automatisch auf der 1. - 3. Ebene statt. Ansonsten wä-re die Darstellung der Verzeichnisstruktur der gesicherten Daten nicht möglich:

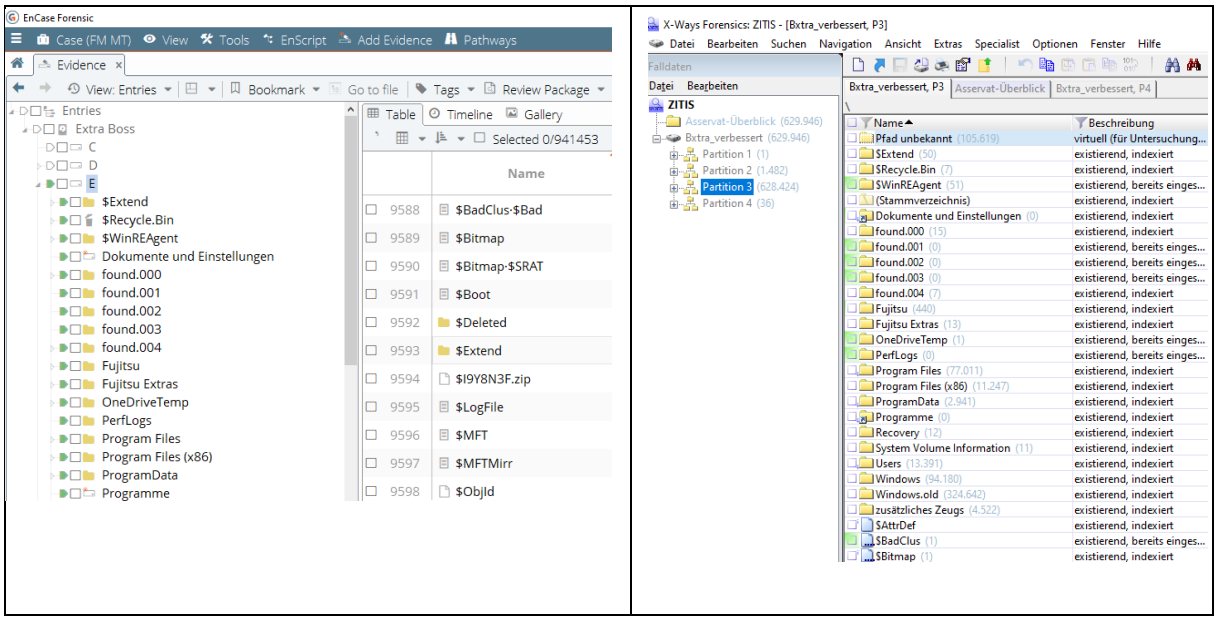


Abbildung 3.2: Verzeichnisstruktur-Darstellung in Encase Forensics 8 und X-Ways Forensics 19

Datenextraktionen auf der 4. Ebene (also z.B. Extraktionen von Compound Files [30]) werden je nach verwendetem Forensik-Tool manuell initiiert (siehe rote Markierungen in den Abbildungen 3.3. und 3.4).

Das gilt auch für die weiteren Verarbeitungsschritte wie z.B. File-Signatur-Analysen oder das Indexieren (siehe orange Markierungen, Abbildungen 3.3 und 3.4):

[30] „Ein sogenanntes Compound File ist eine Datei, die eine Sammlung von Storages und Streams enthält. Ein Compound File besteht aus einem Storage, der beliebig viele Sub-Storages enthalten kann. Jedes Storage kann aus Streams und weiteren Sub-Storages bestehen. Ein Compound File ist daher vergleichbar mit einem Dateisystem: Storages sind Verzeichnisse, Streams sind Dateien.“ Quelle: https://www.it-visions.de /glossar/alle/332/Compound_File.aspx

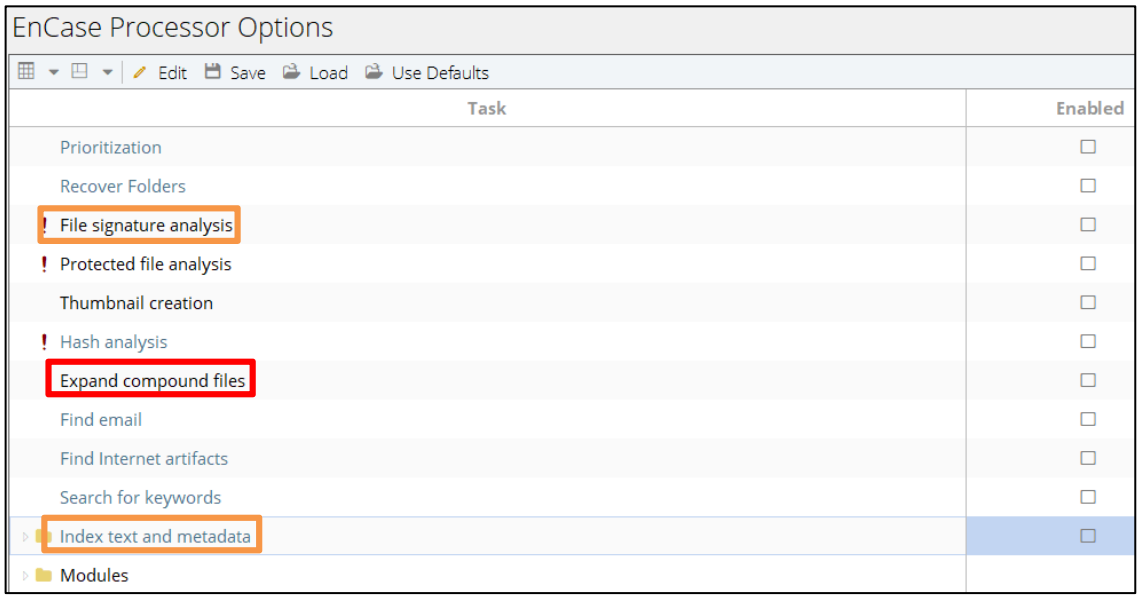


Abbildung 3.3: Einstellungsmöglichkeiten zum Processing in Encase Forensics 8

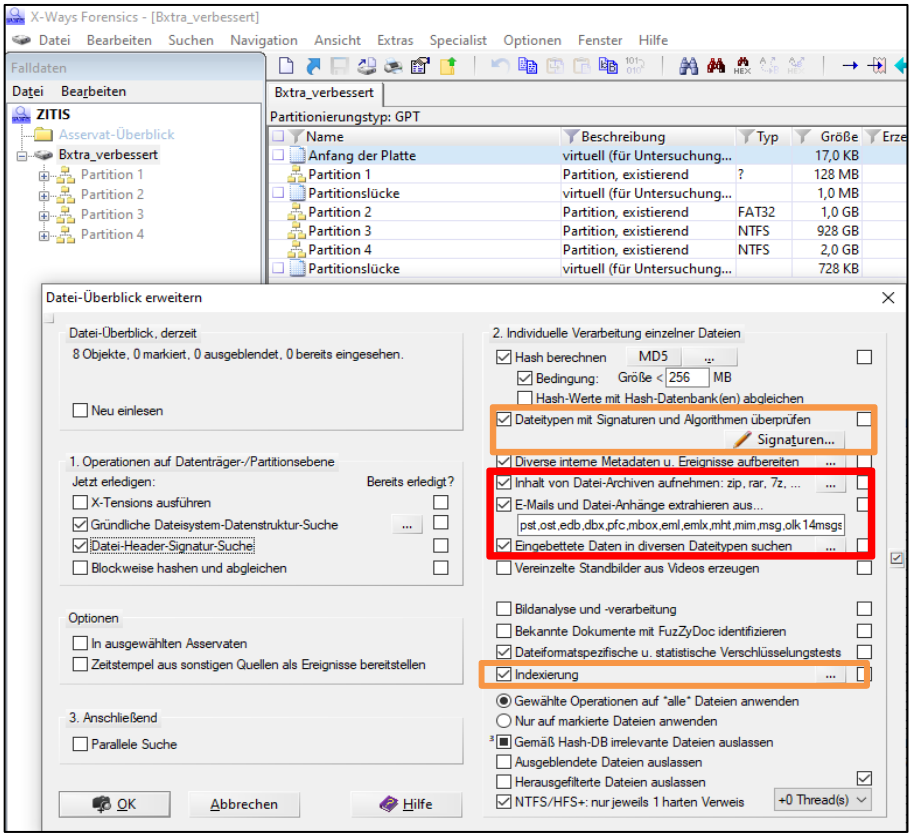


Abbildung 3.4: Konfigurationseinstellungen zur Verarbeitung unter X-Ways Forensics 19

Vergleicht man zusätzlich den Funktionsumfang von den hier eingesetzten eDiscovery- und Forensik-Tools, so ist grundsätzlich (je nach Anwendung) auch hier kein Unterschied feststellbar. Auch das eDiscovery-Tool NUIX kann eine Windows-Registry aufbereiten, gelöschte Dateien wiederherstellen oder Schlupfspeicher extrahieren:

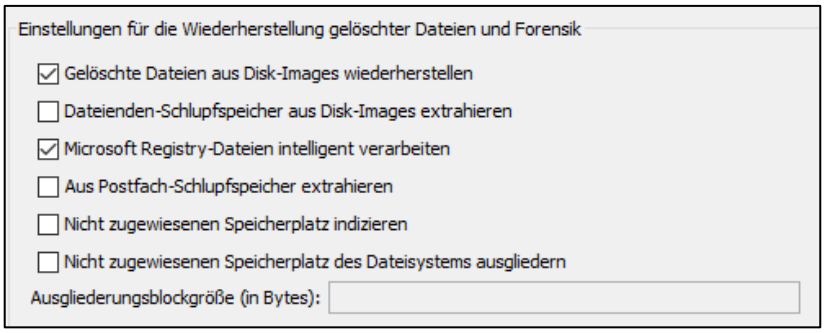


Abbildung 3.5: Einstellungsmöglichkeiten zum Processing unter NUIX WS 8

Ebenso ist die Erweiterung des Funktions- und damit des Verarbeitungsumfanges über Skripte wie bei IT-Forensik- auch bei eDiscovery-Tools gegeben:

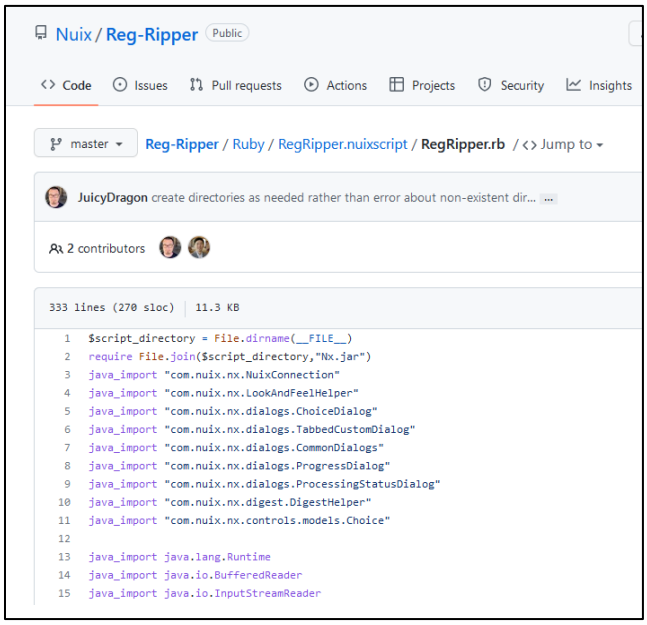


Abbildung 3.6: Ruby-Skript unter NUIX zur Einbindung von RegRipper 3.0

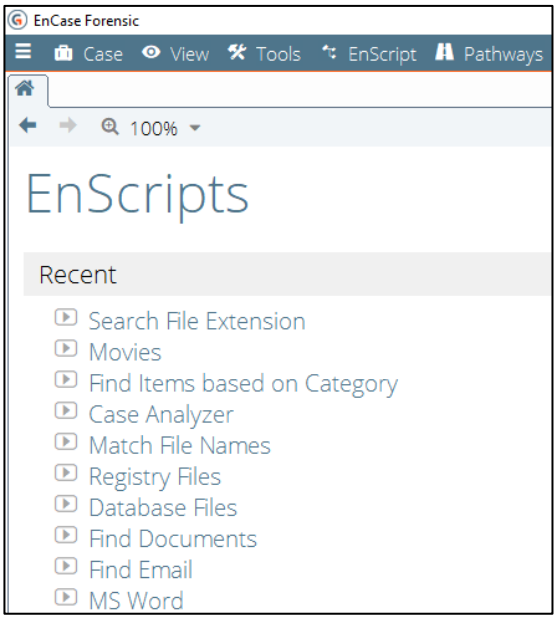


Abbildung 3.7: Enscripts bei Encase Forensics

Verarbeitungs-Ergebnisse im SAP-Modell

In den Anfängen der Entwicklung von IT-Forensik-Tools gab es im Vergleich zu den heutigen Versionen keine vorgefertigten Verzeichnisstrukturen, die je Vorgang automatisch angelegt wurden. Die Ergebnisse zu Analysen oder Suchen wurden im Hauptspeicher (RAM) vorgehalten. Mit dem Anwachsen der Datenmenge und auch der Ausweitung der Analysemöglichkeiten bzw. Funktionsumfänge (z.B. Indexieren) stieg die Menge der Daten an. Daher wurde ein Datenmanagement innerhalb dieser IT-Forensik-Tools nötig, das adäquat mit den extrahierten, gefilterten oder aufbereiteten Analyseergebnissen umgehen konnte. Den Anfang machte der Hersteller Access Data mit seinem Forensik Tool Kit (FTK), indem es sein Programm mit einer Datenbankan-

wendung hinterlegte, in der die Ergebnisse abgespeichert wurden. X-Ways und Encase folgten diesem Lösungsweg der Abspeicherung von Arbeitsergebnissen mit unterschiedlichen Ansätzen wie die Abbildungen 3.8 und 3.9 zeigen:

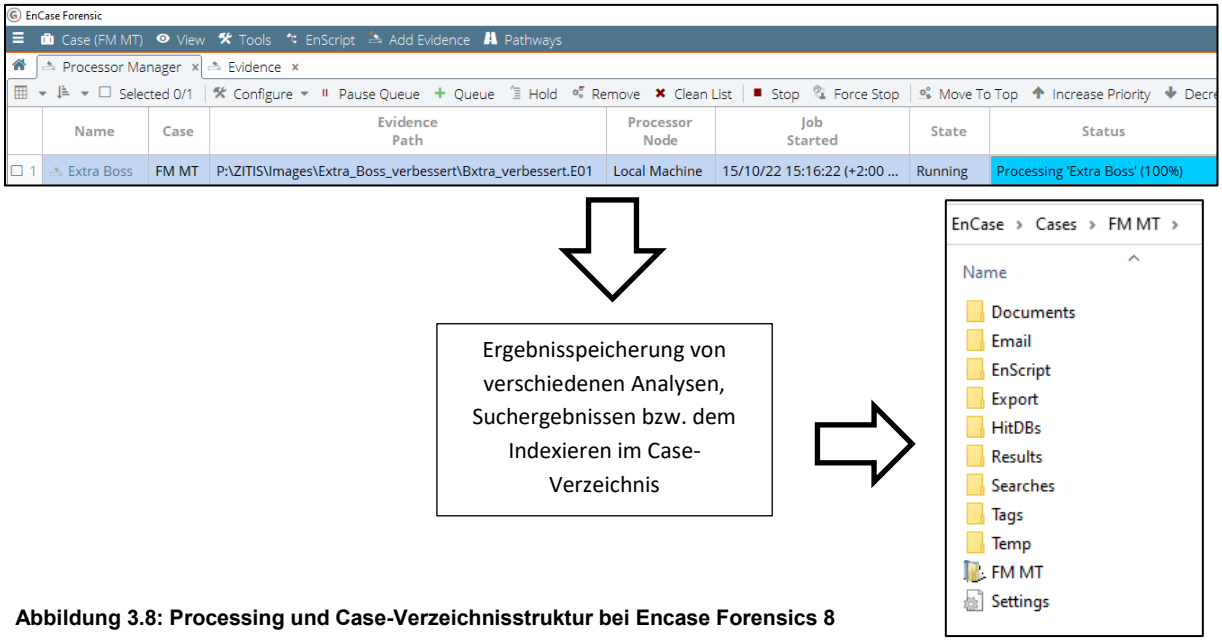


Abbildung 3.8: Processing und Case-Verzeichnisstruktur bei Encase Forensics 8

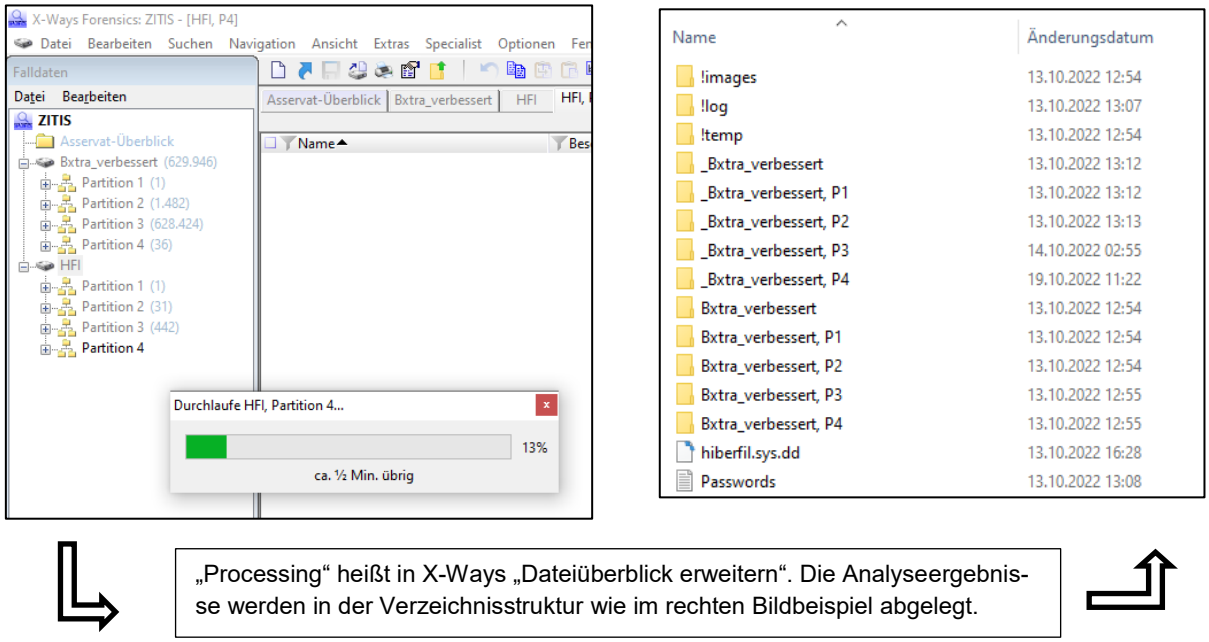


Abbildung 3.9: Processing und Ergebnisabspeicherung in X-Ways Forensics 19.9

Verarbeitungs-Ergebnisse im eDiscovery-Modell

Wogegen in einer eDiscovery-Lösung komplette Datenbanken gebildet werden, in denen dann die Verarbeitungsergebnisse aus dem Extrahieren von Meta- und Inhaltsdaten abgespeichert werden (siehe rot Markierung, Abbildung 3.10):

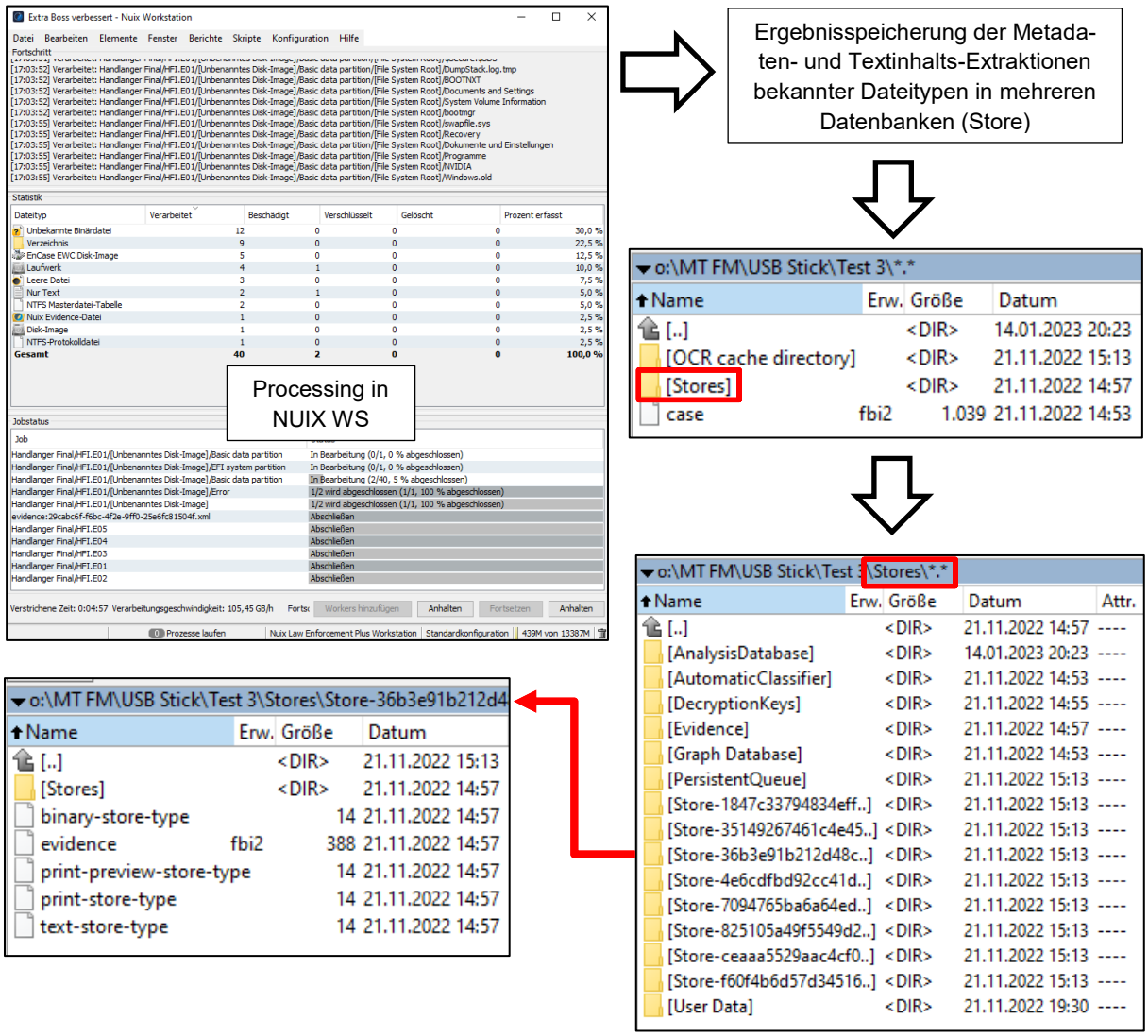


Abbildung 3.10: Processing und Datenbank-Verzeichnisstruktur in NUIX WS 8.8

Um diese Vorgehensweise realisieren zu können, arbeiten eDiscovery-Tools über eine vorgegebene Verzeichnisstruktur, in die die jeweiligen Auslesungsergebnisse von Metadaten und Textinhalten und Rendering-Ergebnisse aus bekannten Datentypen übernommen werden. Das Processing in einem eDiscovery-System ist in der Abbildung 3.11 auf der nächsten Seite vereinfacht dargestellt. Tatsächlich sind eine Vielzahl weiterer Unterprozesse notwendig, um ein Verarbeitungsergebnis zu generieren:

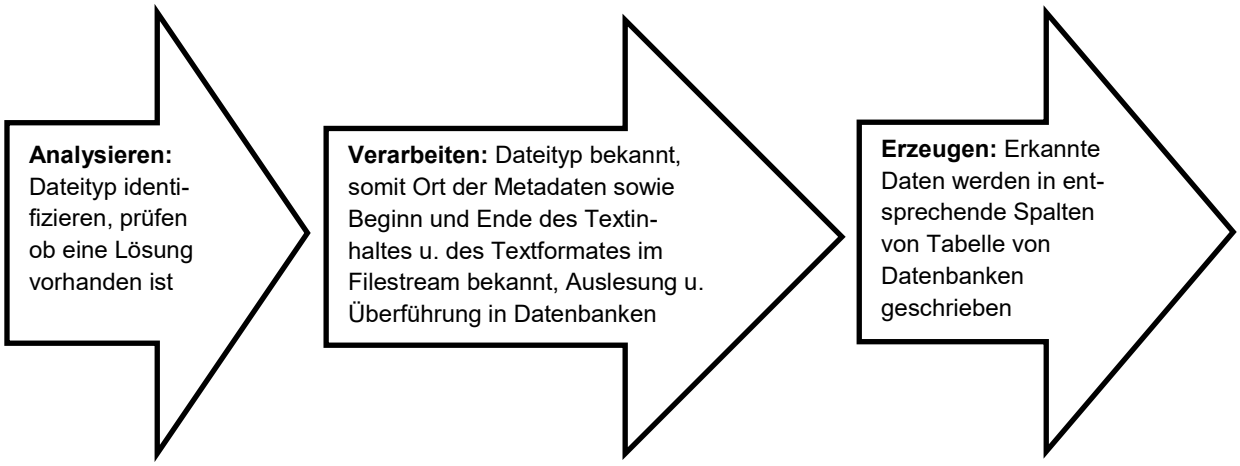


Abbildung 3.11: Workflow in eDiscovery-Systemen

3.1.5 Zusammenfassung und Ergebnis

SAP- und eDiscovery-Modelle haben zwar grundsätzlich die gleiche Zielrichtung, doch die Entwicklung von derartigen Programmen fußt auf unterschiedlichen Ansätzen und führt zu unterschiedlichen Vor- und Nachteilen:

Entwicklung von SAP-Modellen bzw. klassischen IT-Forensik-Programmen

Bei IT-Forensik-Tools ging und geht es in allererster Linie darum Spuren, zu finden und sichtbar zu machen, ohne das originale Spurenmaterial zu verändern, aber dennoch in unterschiedlichste Datenformate Einblick nehmen zu können.

Ursprünglich spielte die „Beherrschung“ großer Datenmengen über ein in der Anwendung integriertes Datenmanagement beim grundlegenden Design von IT-Forensik-Tools keine Rolle. Erst durch neue Funktionen, das Anwachsen der zu untersuchenden Datenmengen und dem Vorhalten von Analyseergebnissen (vormals im Hauptspeicher) wurde die Anforderung notwendig, Aufbereitungen und Analyseergebnisse abzuspeichern. Mittlerweile haben alle namenhaften IT – Forensik-Softwarehersteller Datenbanklösungen in ihre Softwareanwendungen integriert. Zwar lässt sich durch diese Lösung Rechenzeit einsparen, denn bei erneuter Einsichtnahme müssen grundlegende Analysen wie z.B. das Carven, Indexieren oder Recovern nicht wiederholt werden, aber grundsätzlich ist festzustellen, dass bei jedem dieser Tools das Datenmanagement trotz dieser Lösung Grenzen hat. Das macht sich besonders beim Hinzufügen von vielen Images bzw. logischen Datensicherungen bemerkbar. Je nach eingesetztem Forensik-Tool und Hardwareausstattung wird das besser oder schlechter gelöst. Nach Erfahrungen des Autors beginnnen ab ca. 4 - 5 TB verwalteter Daten-

sicherungsgröße die klassischen IT-Forensik Tools instabil oder auffallend langsam zu arbeiten. Entweder dauert es sehr lange bis die Ergebnisse auf dem Monitor nach einer Datenanalyse oder einer Filterung angezeigt werden (Sanduhr läuft) oder die Anwendungen „hängen sich auf“ und nehmen keine Steuerungsbefehle mehr an.

Entwicklung von eDiscovery-Tools

Hier ging es am Anfang darum, eine Softwarelösung zu entwickeln und einzusetzen, bei der die gesetzlichen Vorgaben in US-Zivilprozessen über eine Softwarelösung automatisiert vorbereitet wurden, um den aufwendigen Sichtungs- und Auswerteprozess zur Auffindung relevanter Daten für den Prozess effizienter zu gestalten und so die hohen Kosten für die bisherigen teuren personenbezogenen manuellen Recherchen zu minmieren.

Dieser Lösungsansatz wird über die Harmonisierung der möglichen verfahrensrelevanten Daten und durch Konvertierung in eigene Index-Lösungen erreicht. Derartige Lösungen sind von ihrer Architektur her für die Aufnahme, Verwaltung und Analyse von großen Datenmengen ausgelegt.

Alle Daten mit einer möglichen Verfahrensrelevanz für den eDiscovery-Prozess werden bestimmt, im ersten Verarbeitungsschritt des Cullings unwichtige Dateien abgetrennt und der Rest einer Verarbeitung zugeführt, an deren Ende Informationsinhalte aus den gesicherten Daten über Metadaten- und Inhaltsextraktionen in eine hoch performante Indexlösung überführt werden. Im nachfolgenden Beispiel des Herstellers NUIX wird die Methode dazu vereinfacht dargestellt:

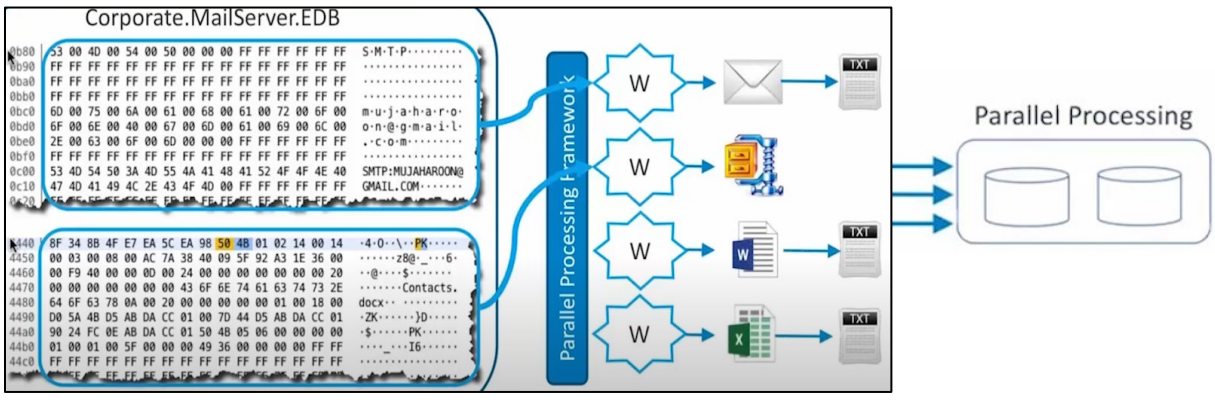


Abbildung 3.12: Auslesen von Metadaten in NUIX

Damit sind eDiscovery-Anwendungen in der Lage sehr große Datenmengen aufzunehmen, ohne dass es zu Performance-Einbrüchen oder Funktionsstörungen wie bei den klassischen IT-Forensik-Tools kommt.

Dadurch wird die Suche und Recherche nach verfahrensrelevanten Daten schnell, einfach und sehr effizient, denn es müssen aus der Gesamtmenge vorhandener Daten nur noch die Suchtreffer auf ihre Geeignetheit für das Gerichtsverfahren überprüft und nicht mehr alle gesicherten Dokumente auf Relevanz gesichtet werden. Dies reduziert den Arbeitseinsatz und finanziellen Aufwand deutlich.
Erst später entstand die Idee, diesen Lösungsansatz auch für IT-forensische Zwecke zu nutzen und Funktionen wie das Extrahieren von forensischen Images, das Carving oder Recovern (Wiederherstellen gelöschter Daten) in den Programmcode zu integrieren. Den Anfang dazu machte das eDiscovery-Tool NUIX, das auch Marktführer in diesem Bereich ist.

Fazit

Die Prozessschritte und Funktionsumfänge sind je nach gewähltem Forensik- bzw. eDiscovery-Tool grundsätzlich gleich, aber unterschiedlich ausgeprägt. Auch NUIX kann carven oder recovern. Auch X-Ways oder Encase können indexieren.

Bei eDiscovery-Tools wird der Verarbeitungsprozess einmal vorkonfiguriert und läuft dann bis zum Ende durch. Bei IT-Forensik-Tools kann jede einzelne Funktion separat aber auch durch Aneinanderreihung, also kombiniert, zum Processing gestartet werden. Auch erneute Analysen mit anderen Zielrichtungen sind nachträglich möglich.

Gravierendster Unterschied zwischen beiden Modellen ist das Auslesen von Meta- und Inhaltsdaten aus bekannten Dateitypen und Kopieren und Indexieren in eigene fallbezogene Verzeichnisse und Index-Dateien.

Damit wird deutlich, dass eDiscovery-Lösungen ihre Stärken in der Verarbeitung und Vorhaltung großer verarbeiteter Datenmengen haben und für bestimmte Dateitypen einen Lösungsansatz besitzen, der automatisiert mit wenig Personalaufwand generiert werden kann. Alle anderen Aufgabenstellungen in der IT-Forensik, die außerhalb dieser vorgefertigten Lösungen stehen, lassen sich jedoch nur mit den klassischen Methoden lösen. Hierzu gehören z.B. die Analyse von Programmen auf ihre Tatrelevanz, der Nachweis von Tathandlungen einer bestimmten Person auf einem PC, der von mehreren Personen genutzt wird und letztendlich wie Software für strafbare Handlungen genutzt wurde (z.B. bei Hacking- oder Kinderpornographie-Verfahren bei der Verbreitung).

In den nun nachfolgenden Testszenarien wird untersucht, wie weit eDiscovery-Lösungen zur Erkennung von digitalen Beweismitteln in Strafverfahren einsetzbar sind und ob diese Einsatzgebiete Risiken im Hinblick auf ihre Beweiskraft beinhalten.

3.2 Tests

3.2.1 Testumgebung

In den nachfolgenden Testszenarien wird mit folgender Testumgebung gearbeitet:

Testgerät Hardware:	Gerätename: DESKTOP-D71OFCB Prozessor: Intel(R) Xeon(R) CPU E5-2660 v3 @ 2.60GHz 2.60 GHz (2 Prozessoren) Installierter RAM: 128 GB (128 GB verwendbar) Geräte-ID: A30271E4-C1EB-4C74-B317-5C65DA56F72E Produkt-ID: 00342-50573-86033-AAOEM Systemtyp: 64-Bit-Betriebssystem, x64-basierter Prozessor
Betriebssystem:	Edition: Windows 10 Pro Version: 21H2 Installiert am: 13.03.2021 Betriebssystembuild: 19044.2006 Leistung: Windows Feature Experience Pack 120.2212.4180.0
eDiscovery-Testsoftware:	**Nuix Workstation** Version 8.8.4.302 MAGNET FORENSICS
Überprüfung der Testergebnisse mit:	GUIDANCE SOFTWARE Program Version Version 8.05.00.182 X-Ways Forensics 19.9 SR-4 x64 (11.02.2020) © 1995-2019 Stefan Fleischmann, X-Ways Software Technology AG Made in Germany. All rights reserved.
Testmaterial: Test 1 Test 2 Test 3:	Lenovo ThinkPad T480 EnCase Evidence File Lenovo ThinkPad T480 EnCase Evidence File USB 8 GB 18.11.2022 13:31 EnCase Evidence

Tabelle 1: Hard- und Software Testumgebung

3.2.2 Test 1 – Verifizierung

Bevor überhaupt eine IT-forensische Untersuchung oder Auswertung erfolgt, werden die beschlagnahmten oder sichergestellten digitalen Spuren- bzw. Datenträger soweit möglich forensisch gesichert. Am Ende des Sicherungsprozesses wird von dem Sicherungsimage, dem gesicherten Verzeichnis oder der Einzeldatei ein Hashwert erzeugt.

Mit den so erzeugten Hashwerten, die im Image selber gespeichert werden, kann später die Originalität bzw. Authentizitität nachgewiesen werden.
Da auch immer wieder Sicherungen aus anderen Behörden zur Untersuchung eingereicht werden, muss deren Unverfälschtheit überprüft werden. Daher wird standardmäßig der bei der Sicherung erstellte Hashwert verifziert, indem bei der Einlesung des Images erneut das Prüfsummenverfahren angewendet und mit den bestehenden Ergebnissen verglichen wird, hier an den Beispielen mit Encase und X-Ways dargestellt:

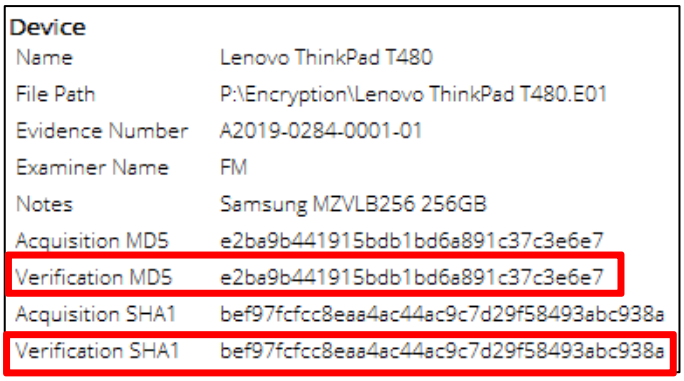


Abbildung 3.13: Verifikations-Beispiel unter Encase Forensics 8

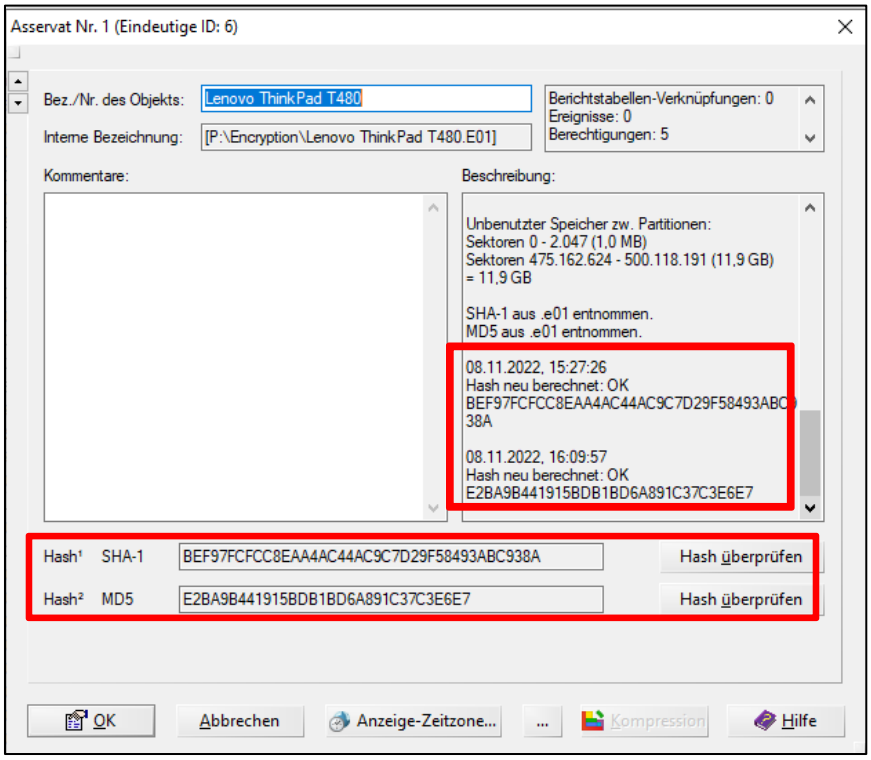


Abbildung 3.14: Verifikations-Beispiel unter X-Ways Forensics 19

Die Tests mit AXIOM und NUIX sollen nun zeigen, ob diese Verifikation bei der Einlesung auch durchgeführt oder zumindest diese Funktion angeboten und protokolliert wird.

Test mit AXIOM 6.5:

AXIOM bietet keine Funktion zur Überprüfung fremder Prüfsummen aus deren forensischen Images an. Weder im Tool selber, noch in der Bedienungsanleitung oder im Internet fanden sich Hinweise auf diese Möglichkeit.

AXIOM kann digitale Datenträger mit einer eigenen Funktion sichern und dann eigene Hashwerte berechnen. Dazu muss in der Konfiguration diese Funktion vor dem Verarbeitungsprozess aktiviert werden (siehe Abbildung 3.15 rechts).

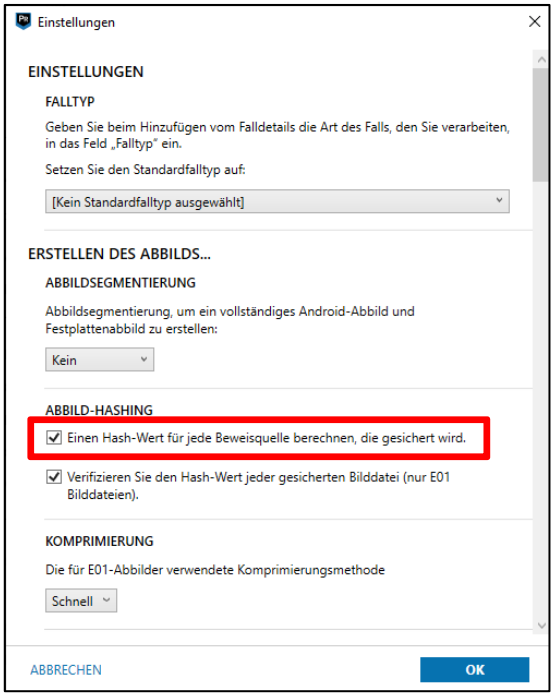


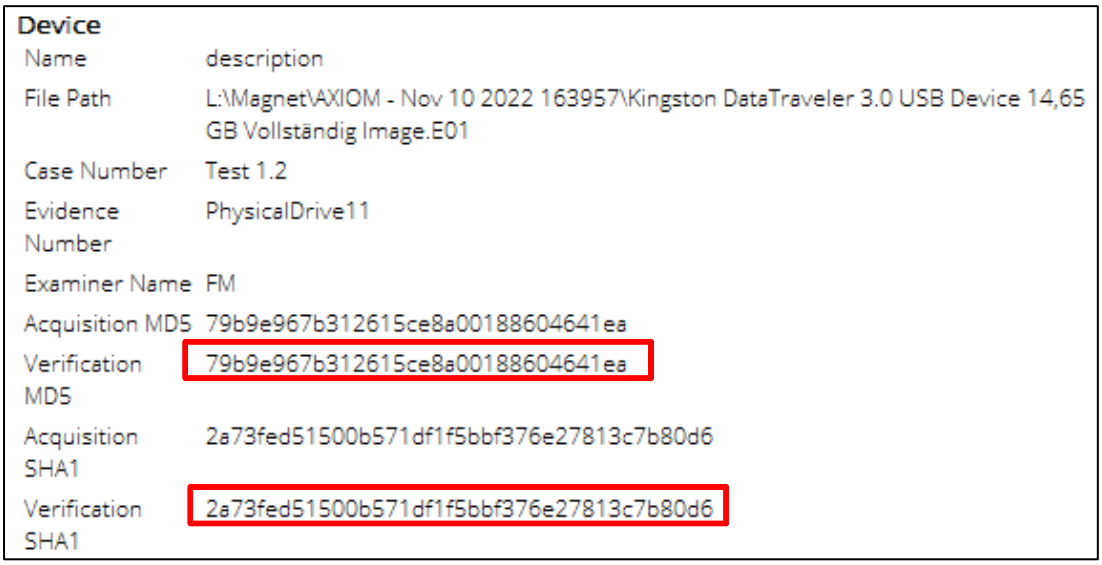
Abbildung 3.15: Konfigurationsmöglichkeiten zum Hashen unter AXIOM

Über einen Test mit einem Speichermedium (USB Stick 3.0 Kingston DataTraveler) wurde diese Funktion nachvollzogen (siehe nächste Abbildung 3.16):

```
Imager Product: Magnet AXIOM Process
Imager Version: 6.5.0.32778

Examiner Name: FM
Evidence Number: Kingston DataTraveler 3.0 USB Device 14,65 GB Vollständig Image
Description:

Relative Activity Log Path: activity_log.txt
Original Activity Log Path: L:\Magnet\AXIOM - Nov 10 2022 163957\activity_log.txt
Activity Log MD5 Hash: A116825FF388A80B7DC6CB8CD3E2C201

Output Directory: AXIOM - Nov 10 2022 163957
Full Output Directory: L:\Magnet\AXIOM - Nov 10 2022 163957

Total Segments: 1

Relative Segment 1 Path: Kingston DataTraveler 3.0 USB Device 14,65 GB Vollständig Image.E01
Full Segment 1 Path: L:\Magnet\AXIOM - Nov 10 2022 163957\Kingston DataTraveler 3.0 USB Device 14,65 GB Vollständig Image.E01
Segment 1 MD5 Hash: 79b9e967b312615ce8a00188604641ea
Segment 1 SHA1 Hash: 2a73fed51500b571df1f5bbf376e27813c7b80d6
```

Abbildung 3.16: Protokolldatei *image_info.txt* in AXIOM

Doch das Programm ist nicht in der Lage, die Hashwerte aus fremden Sicherungsimages zu überprüfen. Diese Verifikations-Funktion ist in AXIOM nicht enhalten. Mit Encase wird überprüft, ob diese Hashwerte korrekt gebildet werden:



Device	
Name	description
File Path	L:\Magnet\AXIOM - Nov 10 2022 163957\Kingston DataTraveler 3.0 USB Device 14,65 GB Vollständig Image.E01
Case Number	Test 1.2
Evidence Number	PhysicalDrive11
Examiner Name	FM
Acquisition MD5	79b9e967b312615ce8a00188604641ea
Verification MD5	79b9e967b312615ce8a00188604641ea
Acquisition SHA1	2a73fed51500b571df1f5bbf376e27813c7b80d6
Verification SHA1	2a73fed51500b571df1f5bbf376e27813c7b80d6

Abbildung 3.17: Überprüfung der Prüfsummenwerte aus AXIOM mit Encase Forensics 8.05

Test mit NUIX Workstation 8.8

Das Tool enthält ebenfalls keine Verifizierungsmöglichkeit. NUIX WS 8.8 fertigt zwar bei jedem Verarbeitungsprozess automatisch MD5-Hashwerte von jeder Einzeldatei an und bietet optional auch SHA-1-Prüfsummenbildungen über die Vorkonfiguration zum Processing an, doch es wird genauso wie bei AXIOM keine Verifizierungsfunktion für einzulesende und zu verarbeitende forensische Images angeboten.

Über eine Internetrecherche wurde herausgefunden, dass ebenso wie bei AXIOM auch NUIX über einen zusätzlichen einfachen und einen umfangreichen Enterprise-(Unternehmens)Imager verfügt, der die erstellten forensischen Images ebenfalls mit Hashwerten versehen kann.

Wegen fehlender Lizenzen konnte nicht geprüft werden, ob diese Tools über Verifizierungsmöglichkeiten verfügen. Bei Youtube wurden Videoclips zum Tool NUIX Enterprise Collection Center ausgewertet – auch aus dieser Auswertung ergaben sich keine Hinweise auf eine Verifikationsmöglichkeit. Das Design dieses Enterprise-Imager-Tools unterstützt die Annahme des Autors, dass bei dem ursprünglichen Einsatzgebiet von eDiscovery-Tools nur von Datensammlungen innerhalb von Firmen ausgegangen wurde und nicht von außerhalb eines Unternehmens Daten einem Verarbeitungsprozess zugeführt wurden. Wie man aus der Abbildung 3.18 sehen kann, wird das Tool in die Firmendomäne integriert und kann so über das Netzwerk Verbindungen zu den einzelnen Clients aufbauen und Images erstellen. Genauso wie bei AXIOM und den

klassischen Forensik-Tools können dann hiervon Prüfsummen erzeugt werden:

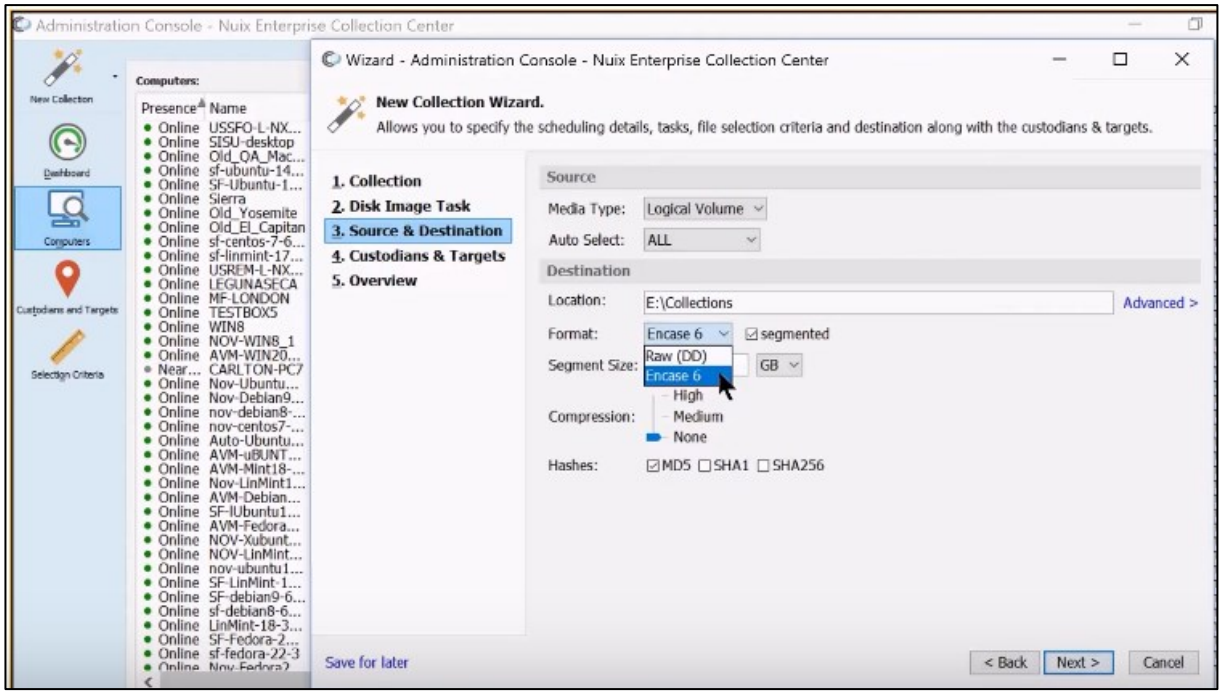


Abbildung 3.18: Erweiteter Imager - NUIX Enterprise Collection Center

Zusammenfassung Test 1

Beide getestete eDiscovery-Tools verfügen über keine eigene Verfizierungsfunktion zur Überprüfung von Hashwerten aus forensischen Images, die mit anderen Tools hergestellt wurden und durch diese eDiscovery-Tools verarbeitet werden sollen. Es wurden keine Leitlinien oder sonstigen Dokumente im EDRM Gremium gefunden, die eine Verifizierung vorschreiben bzw. empfehlen. Als Ursache wird hierfür die Tatsache angenommen, dass die Daten für Zivilklageverfahen nur aus der eigenen Firma erhoben werden. Aus Sicht des Autors birgt diese fehlende Funktionalität ein gewisses Risiko in Strafverfahren. Sollte die Verifizierung über andere Forensik-Tools vergessen werden, bietet dies Anlass, die Authentizität eingelesener Sicherungen zumindest in Frage zu stellen.

In den USA ist der Nachweis der geschlossenen Beweismittelkette [31] (chain of custody) in Strafverfahren von großer Bedeutung und kann bei nachgewiesenen Zweifeln an der Authentizität zum Ausschluß dieses Beweises führen. Spätestens wenn in deutschen Strafverfahren nachträglich dieser Verifzierungsprozess eingefordert wird und das Negativergebnis nicht erklärbar ist, bietet dies auch hier erheblichen Anlass die gefundenen Spuren in Zweifel zu ziehen. Es besteht dann der Verdacht der Be-

[31] Quelle: https://it-forensik.fiw.hs-wismar.de/index.php/Chain_of_Custody

weismittelverfälschung. Anders als im anglo-amerikanischen Strafrecht hat aber an deutschen Gerichten der urteilende Richter das Recht auf freie Beweiswürdigung gem. § 261 StPO, so dass nicht automatisch wie im US-Strafrecht ein Beweismittelverwertungsverbot droht.

3.2.3 Test 2 – Unbekannte Dateitypen in eDiscovery-Systemen

Während klassische forensische Tools grundsätzlich alle Dateien einer Sicherung, unabhängig von deren Dateiformat anzeigen, kann ein eDiscovery-System nur die Daten nach einem Verarbeitungsprozess darstellen, deren Format es erkennt und dafür eine Verarbeitungs- und Darstellungslösung besitzt.
Es besteht also aufgrund der Programmarchitektur in eDiscovery-Anwendungen grundsätzlich das Risiko, dass Dateien, die eine Relevanz für das Strafverfahren besitzen, aber nicht von den Dateiformat-Lösungen des eDiscovery-Herstellers unterstützt bzw. erkannt werden, durch den Datenaufbereitungsprozess gar nicht in das Endprodukt, die Index-Lösung, „geschrieben“ und somit in der anschließenden Darstellung auch nicht angezeigt werden bzw. recherchierbar sind.
Der einzige Leitfaden, welcher sich im EDRM-Gremium mit dieser Thematik beschäftigt, befindet sich im Dokument *5.0 Reporting* [32] unter *5.5 Ausnahmeberichte*:

„Dateien, die nicht verarbeitet werden können, sollten in der verarbeitenden Datenbank als Ausnahme gekennzeichnet werden. Dies sind Dateien, für die kein Text oder Metadaten extrahiert werden können oder für die keine Abbildung gerendert werden kann. Diese Kategorie kann verschlüsselte oder beschädigte Dateien, Systemdateien, Programmdateien oder andere Arten umfassen, die keine Informationen darstellen.
Ausnahmeinformationen sollten für die Suche und Berichterstellung verfügbar sein. Im Idealfall liefert der Bericht den Grund, warum die Dateien nicht verarbeitet werden konnten. Ein Ausnahmebericht kann beispielsweise die folgenden Informationen enthalten:
Dateiname, ursprünglicher Verzeichnisspeicherort der Datei, Grund für die Ausnahme (Fehler). Gründe für Dateiausnahmen können Dateibeschädigung, Verschlüsselung, Kennwortschutz, Virusinfektion, Null-Byte-Datei oder NIST-Ausschluss sein.“

Damit wird deutlich: es wird keine eindeutige Protokollierung für nicht unterstützte Dateitypen verlangt. Es werden lediglich einige Beispiele wie Verschlüsselung oder Dateibeschädigung als Kann-Regelung erwähnt, nicht unterstützte Dateitypen werden ausdrücklich nicht erwähnt. Dieser Test 2 untersucht daher, wie die eingesetzten eDiscovery-Tools NUIX und AXIOM mit nicht bekannten Datentypen umgehen.

[32] https://edrm.net/wiki/5-0-reporting/

Testdaten – Untersuchungsergebnis aus der klassischen Untersuchung

Als Testobjekt dient ein forensisches Image aus dem beruflichen Hintergrund des Autors. Der damalige Untersuchungsauftrag bestand darin, im Sicherungsimage Dokumente zu finden, die den Verdacht erhärten, dass der Beschuldigte aus seiner Rolle als öffentlich bestellter Kfz-Prüfingenieur und TÜV-Gutachter falsche amtliche Gutachten (Hauptuntersuchung – Zulassung für den Strassenverkehr) mit seinem dienstlichen Notebook erstellt hatte. Im ersten Untersuchungsansatz konnten über die üblichen Filterfunktionen (Dokumente und Druckerzeugnisse) und manuelle Suche im Image keine belastbaren Dokumente gefunden werden. Auch ein “recovern” bzw. “carven” brachte keinen Erfolg.

Auf Nachfrage zur Arbeitsweise bei der Erstellung von Gutachten beim Arbeitgeber wurde erklärt, dass dies durch die Nutzung einer eigenen Datenbankanwendung erfolgte. Daraufhin wurde die identifizierte Nutzdatei ZEUS_Working.mdf aus der Sicherung extrahiert und in einer Virtuellen Maschine in eine Laufzeitumgebung gebracht, um so Einblick zu bekommen (siehe Abbildungen 3.19 und 3.20) und Einträge aus dem tatrelevanten Zeitraum näher zu untersuchen:

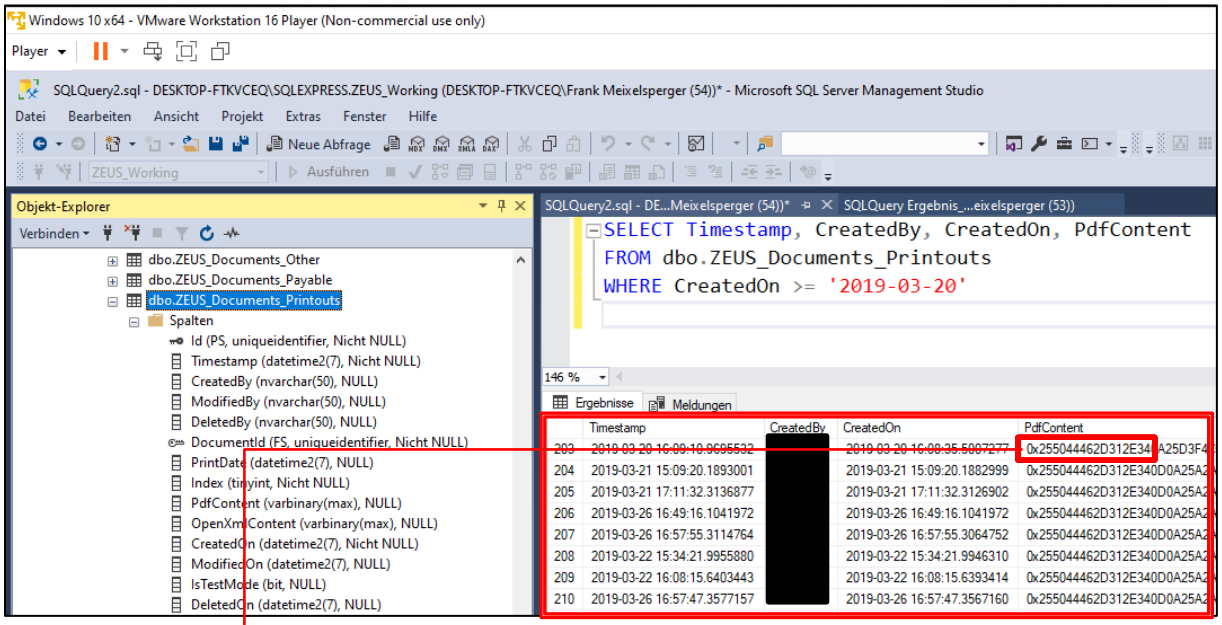


Abbildung 3.19: Einsichtnahme und Suche in ZEUS_working.mdf

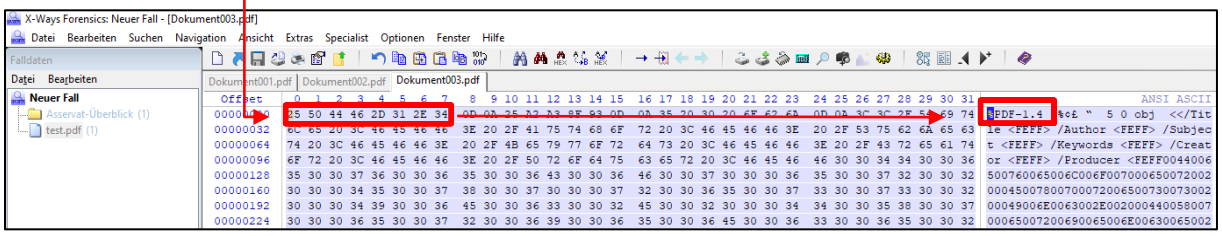
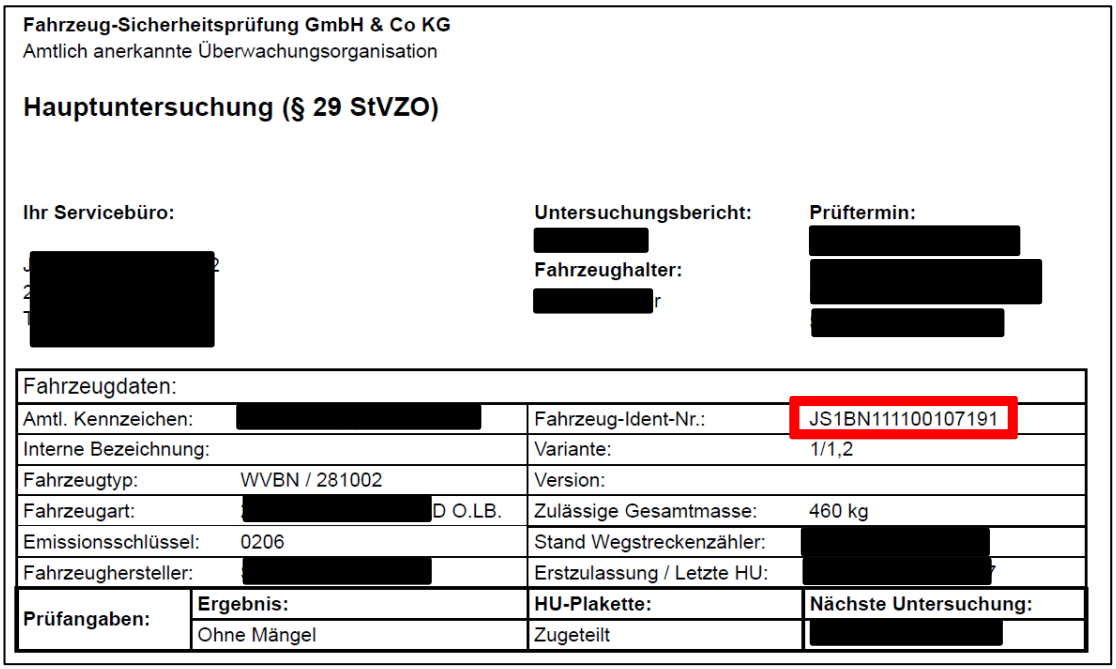


Abbildung 3.20: Abspeicherung eines PDF Filestreams in einer leeren Datei

Durch die Einsichtnahme wurde die Abspeicherung von PDF-Dokumenten als Filestreams innerhalb einer Datenbank erkannt. Damit war es möglich ca. 200 verfahrensrelevante PDF-Filestreams innerhalb der Datenbank zu finden und mittels X-Ways Forensics wieder sichtbar zu machen, indem diese Filestreams in eine leere Datei kopiert und mit der Dateiendung *.pdf abgespeichert wurden (siehe Abbildung 3.21).
Die rot markierte Zahlenreihe ist ein Suchbegriff, der zur Suche nach verfahrensrelevanten Treffern in den jeweiligen Verarbeitungsergebnissen von AXIOM und NUIX später genutzt wurde. Da es sich um eine Individualnummer handelt, eignet sie sich besonders gut dafür. Ein Suchtreffer dürfte nur einmal als Ergebnis erscheinen. Alle anderen Suchfelder sind aus datenschutzrechtlichen Gründen abgeblendet:



Fahrzeug-Sicherheitsprüfung GmbH & Co KG
Amtlich anerkannte Überwachungsorganisation

Hauptuntersuchung (§ 29 StVZO)

Ihr Servicebüro: [redacted]

Untersuchungsbericht: [redacted]
Fahrzeughalter: [redacted]

Prüftermin: [redacted]

Fahrzeugdaten:			
Amtl. Kennzeichen:	[redacted]	Fahrzeug-Ident-Nr.:	JS1BN111100107191
Interne Bezeichnung:		Variante:	1/1,2
Fahrzeugtyp:	WVBN / 281002	Version:	
Fahrzeugart:	[redacted] D O.LB.	Zulässige Gesamtmasse:	460 kg
Emissionsschlüssel:	0206	Stand Wegstreckenzähler:	[redacted]
Fahrzeughersteller:	[redacted]	Erstzulassung / Letzte HU:	[redacted]

Prüfangaben:	Ergebnis:	HU-Plakette:	Nächste Untersuchung:
	Ohne Mängel	Zugeteilt	[redacted]

Abbildung 3.21: Sichtbar gemachtes PDF-Dokument aus einem Filestream der ZEUS_working.mdf

Der aufgestellten These folgend, dürften diesen PDF-Dokumente im Verarbeitungsergebnis von eDiscovery-Programmen wie AXIOM und NUIX nicht angezeigt werden, da diese im Analyseprozess vor der Verarbeitung einen MDF-Dateityp nicht erkennen und somit für die o.g. verfahrensrelevante Datenbankdatei ZEUS_working.mdf keine Verarbeitungslösung besitzen.
Damit kann die enthaltene aus verknüpften Tabellen bestehende Verzeichnisstruktur dieser Datenbankdatei nicht extrahiert und auch nicht die in den einzelnen Tabellen enthaltenen PDF-Filestreams über eine File-Type-Analyse von Header-Daten, erkannt werden.

Test mit MAGNET AXIOM 6.5

Konfiguration zum Processing

AXIOM unterstützt einen Umfang von ca. 100 Dateitypen, wozu auch PDF – Dokumente gehören, bietet aber laut herstellereigener Bedienungsanleitung nur für SQLite-Datenbanken Unterstützung [33]. SQLite-Datenbanken werden häufig für Instant Messenger wie Threema, Signal oder WhatsApp eingesetzt. SQL Server Master Datenbankdateien (*.mdf) werden jedoch nicht unterstützt.

Obwohl die Magnet Forensics Dokumentation zum AXIOM-Tool die Aufbereitung derartiger Datenbankdateitypen verneint, versucht der Autor über die Hinzufügung einer PDF-Header-Information in der Custom File Type Liste dieses Programms das Defizit zu kompensieren (siehe rote Markierung in Abbildung 3.22):

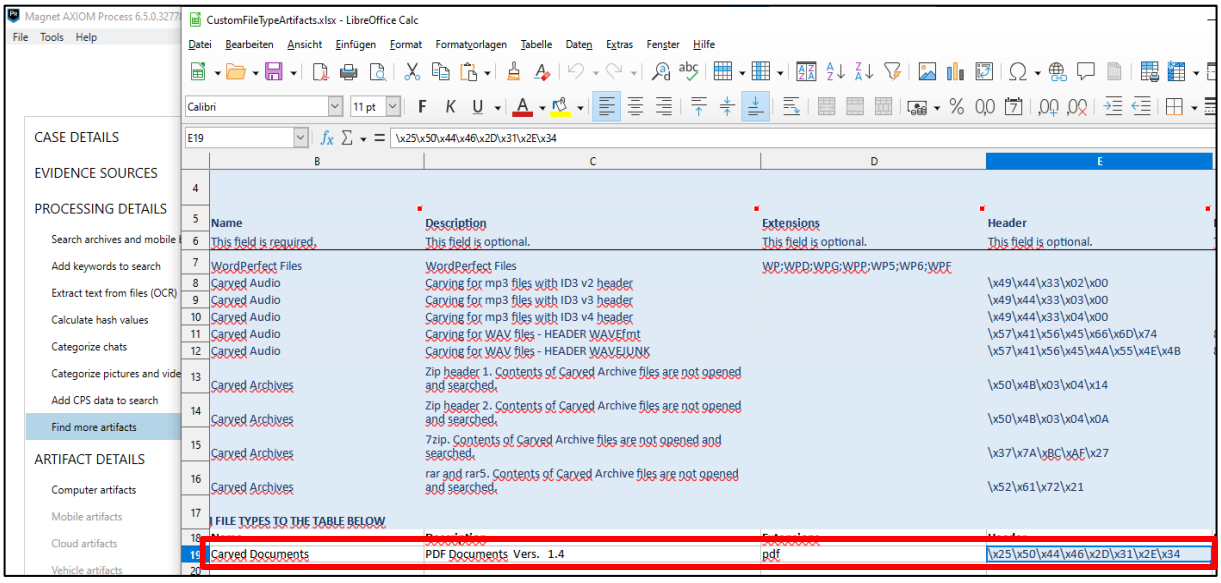


Abbildung 3.22: Erzeugung eines neuen Headers, hier für PDF Dokumente

und ergänzt für den Erkennungs- und Verarbeitungsprozess die Konfiguration um einen Header-Neueintrag (siehe rote Markierung, Abbildung 3.23):

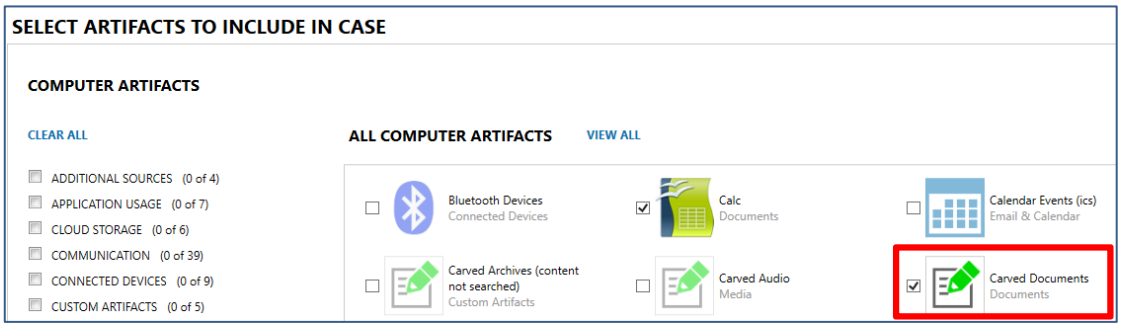


Abbildung 3.23: Aktivierung der Carving Suche nach PDF-Dokumenten

[33] https://www.magnetforensics.com/docs/axiom/html/Content/en-us/axiom/processing-evidence/processing-daf.htm?tocpath=Configuring%20processing%20details%7C_____8

Nach der Verarbeitung über die Anwendung AXIOM-Process wird das Ergebnis in der Anwendung AXIOM Examine angezeigt. Ergebnis: Matching Result (0 of 0).

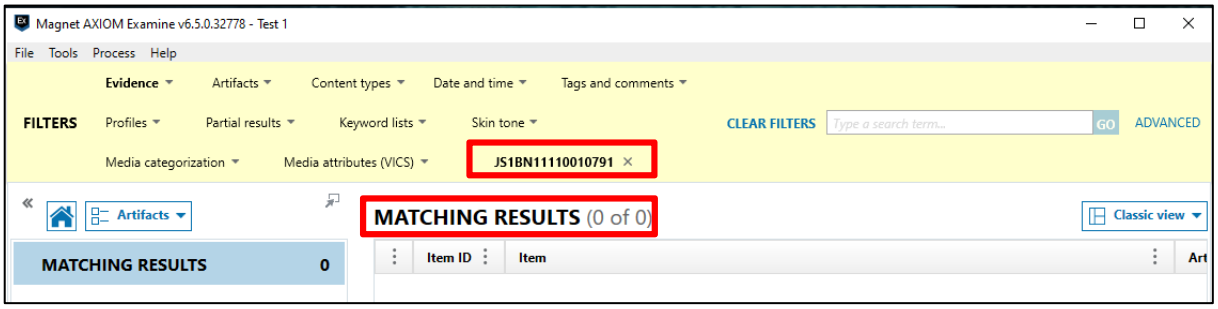


Abbildung 3.24: Treffergebnis für Suchbegriff "JS1BN11110010791" in AXOIM

Zusätzlich wird über eine Individualnummer (siehe hierzu Abbildung 3.24, rote Markierung) explizit im Datenbestand ein weiterer Versuch unternommen – mit gleichem Resultat.

Ein weiterer Versuch wird unternommen, indem über die Änderung der Darstellungsansicht auf "Dateisystem" manuell nach dem bekannten Ablageort der verfahrensrelevanten Datenbankdatei namens *ZEUS_Working.mdf* gesucht wird. Dazu wird die ZEUS_Working.mdf markiert und in der HEX-Anzeige nach dem Individualbegriff gesucht – ebenso ohne Erfolg (siehe dazu Abbildung 3.25):

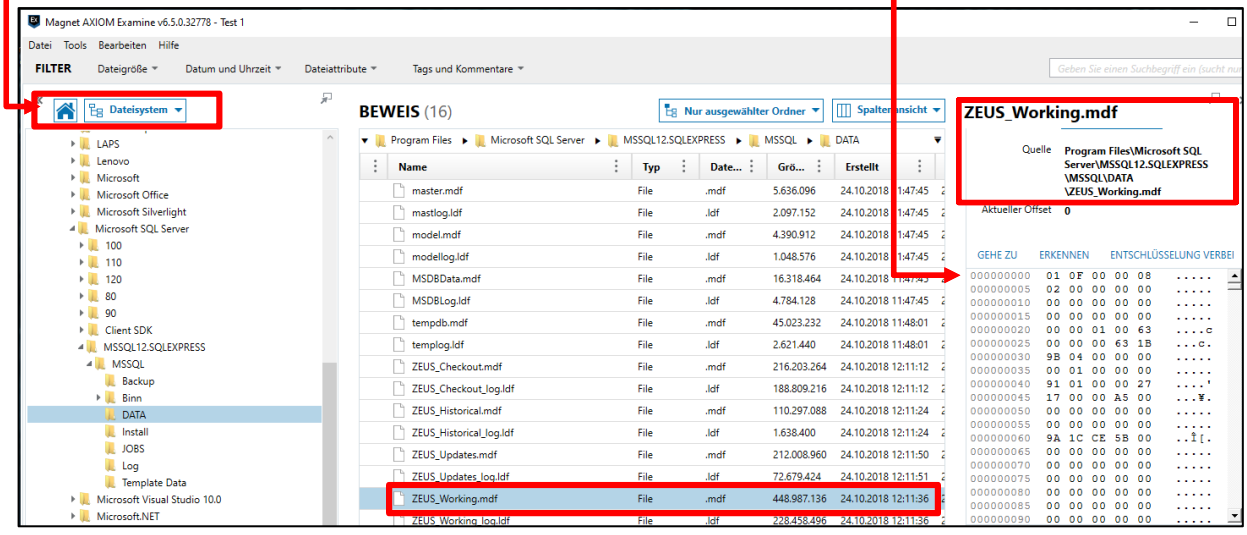


Abbildung 3.25: Dateisystem-Ansicht unter AXIOM

Untersuchung der Protokolldaten

Abschließend wird untersucht, wie AXIOM mit nicht erkannten bzw. verarbeitbaren Datentypen umgeht, also in welcher Form eine Protokollierung hierzu stattfindet. AXIOM bietet die Möglichkeit an, die Protokollsammlung in einer ZIP – Datei abzu-

speichern. Innerhalb dieser Sammlung verschiedener TXT-Dateien wurde dann nach Einträgen bzw. Hinweisen für nicht verarbeitete Datentypen gesucht – ohne Erfolg (siehe Abbildung 3.26)

Eine Nachfrage beim Support AXIOM Deutschland ergab, dass nicht verarbeitete Dateitypen in der Protokolldatei *artifacts.log* festgehalten werden sollen. Doch auch hier konnte kein Hinweis auf die Nichtverarbeitung einer MDF-Datei festgestellt werden (siehe Abbildung 3.26). Weder der gesuchte Datenbankdateityp *.mdf noch andere nicht verarbeitete Dateitypen wurden in diesem Protokoll abgespeichert. Hier konnten lediglich die Ergebnisse verarbeitbarer Datentypen festgestellt werden:

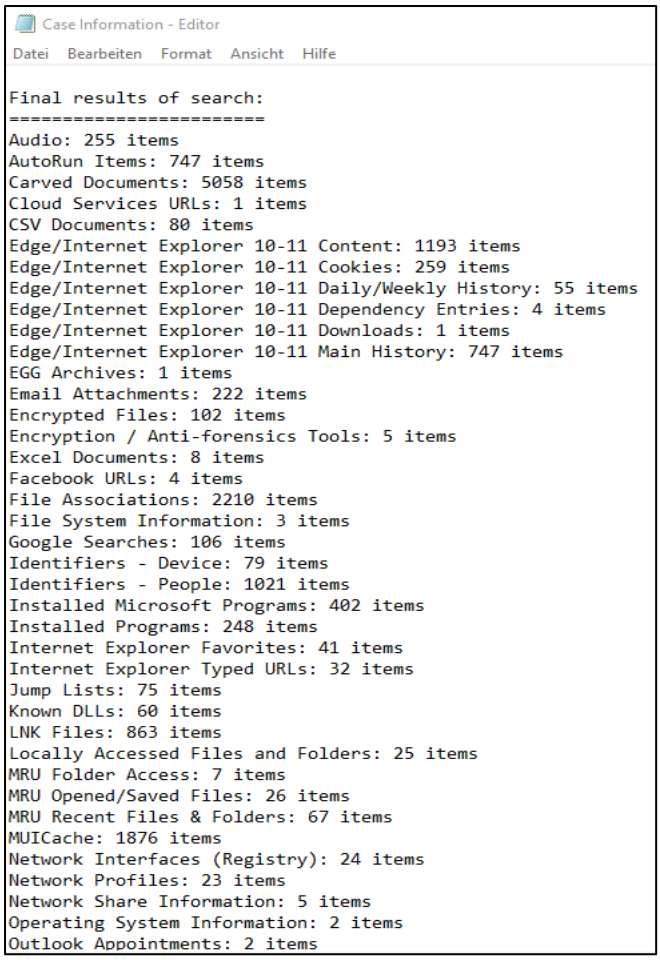


Abbildung 3.26: Auszug aus der Protokolldatei *case Information.log* in AXIOM

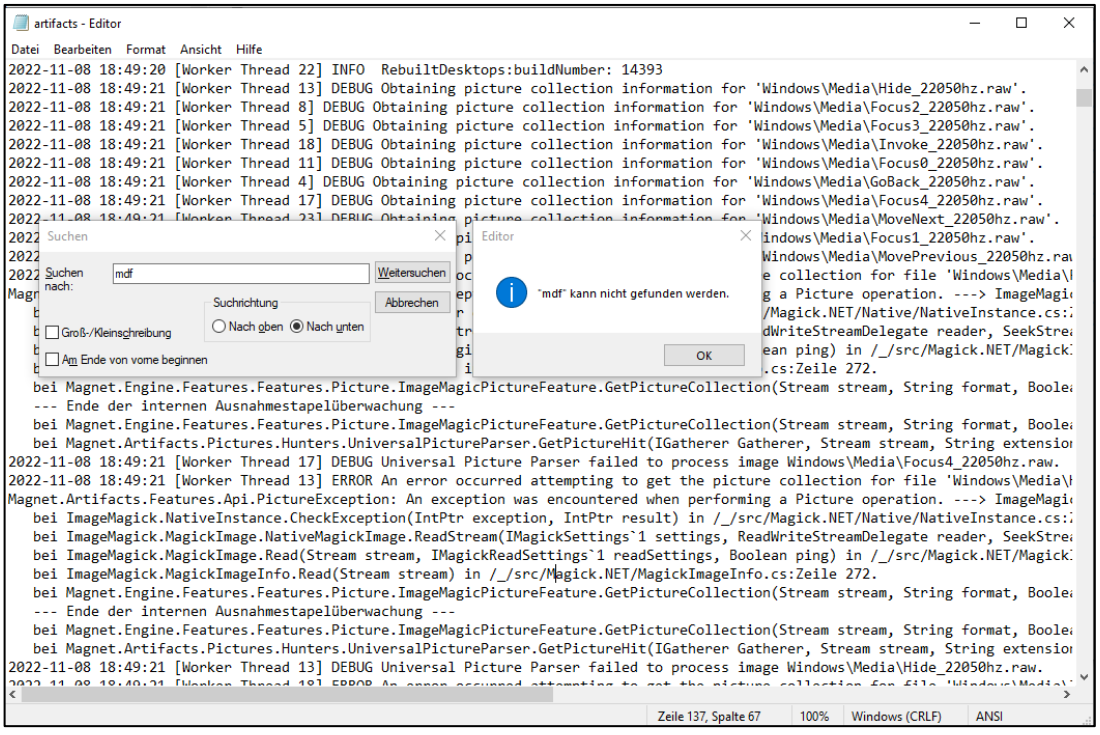


Abbildung 3.27: Protokolldatei artifacts.log in AXIOM

Test mit NUIX Workstation 8.8

Vorbemerkungen

Der Softwareanbieter NUIX hält für seine Anwendung *NUIX Workstation* verschiedene Support-Dokumente (siehe Abbildung 3.28) vor. Laut diesem Dokument *Supported File Types* werden keine SQL-Datenbank-Dateien des Typs MDF unterstützt:

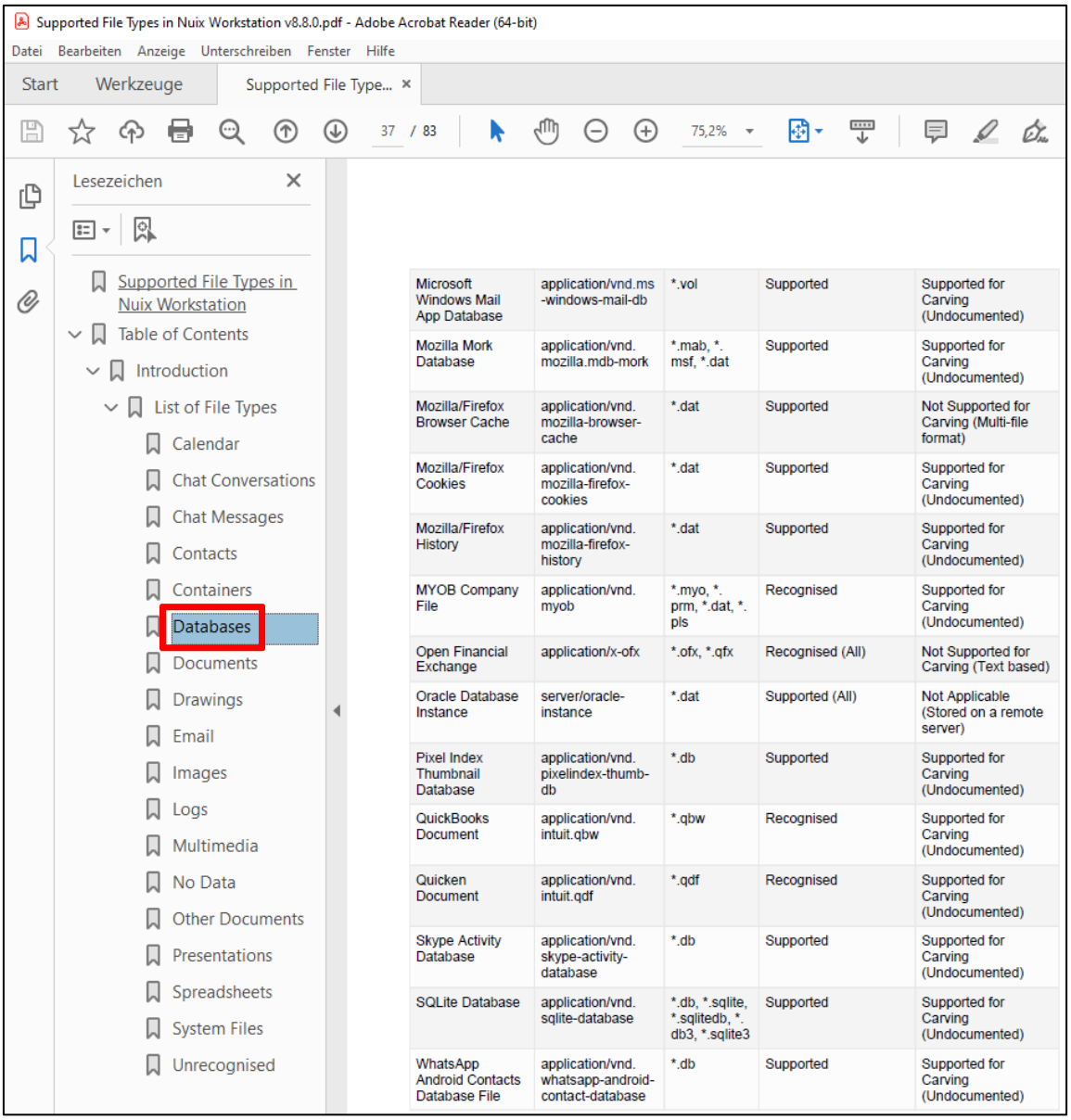


Abbildung 3.28: NUIX Workstation – Supported File Types

Trotzdem wird ein Testlauf durchgeführt, bei dem alle Möglichkeiten zur Extraktion von Datenbank-Dateien ausgeschöpft werden.

Konfiguration zum Processing unter NUIX WS 8.8

In dieser Konfigurationsmaske (Evidenz-Verarbeitungseinstellungen) wird die Erkennung und Verarbeitung von Datenbanken, also strukturierten Massendaten, aktiviert.

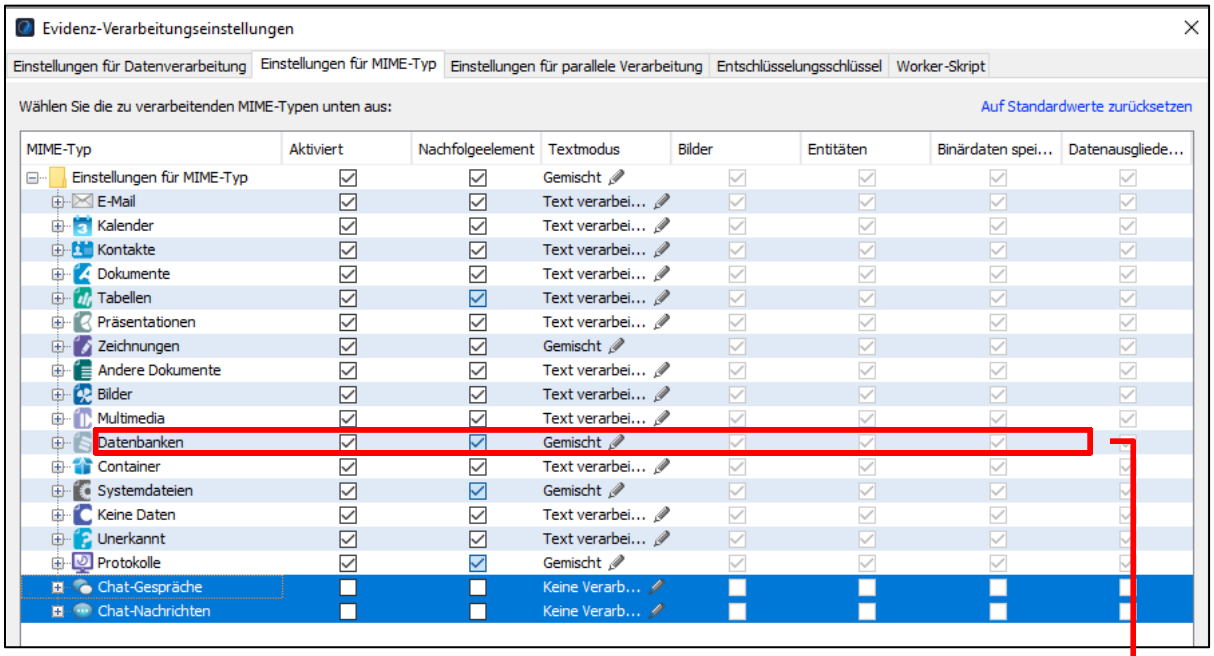


Abbildung 3.29: NUIX Workstation MIME-Type Einstellungen

Durch die Öffnung des Unterverzeichnisses von Datenbanken wird der Umfang unterstützter Datenbankformate deutlich. MDF – Dateitypen befinden sich nicht darunter, wie bereits im o.g. Dokument (supported File Types.pdf) festgestellt wurde.

Trotzdem wird ein Test mit artverwandten Einstellungen versucht. Als Testimage dient wieder das unter AXIOM bereits eingesetzte TÜV-Image (Lenovo Thinkpad T480.E01):

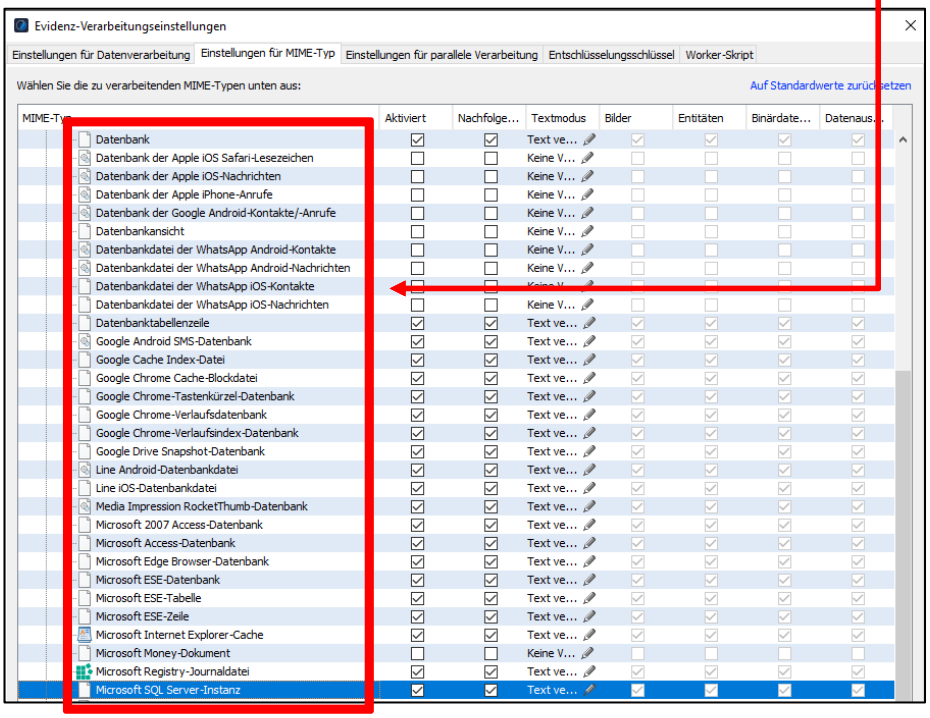


Abbildung 3.30: Unterverzeichnis der MIME-Types - Bereich Datenbanken

Wie erwartet, werden keine Verarbeitungsergebnisse für den Dateityp *.mdf angezeigt. Auch die stichpunktartige Einsichtnahme in die vorhandenen Arbeitsergebnisse unter Datenbanken (siehe rote Markierung in Abbildung 3.31) führte zu keinem Erfolg.

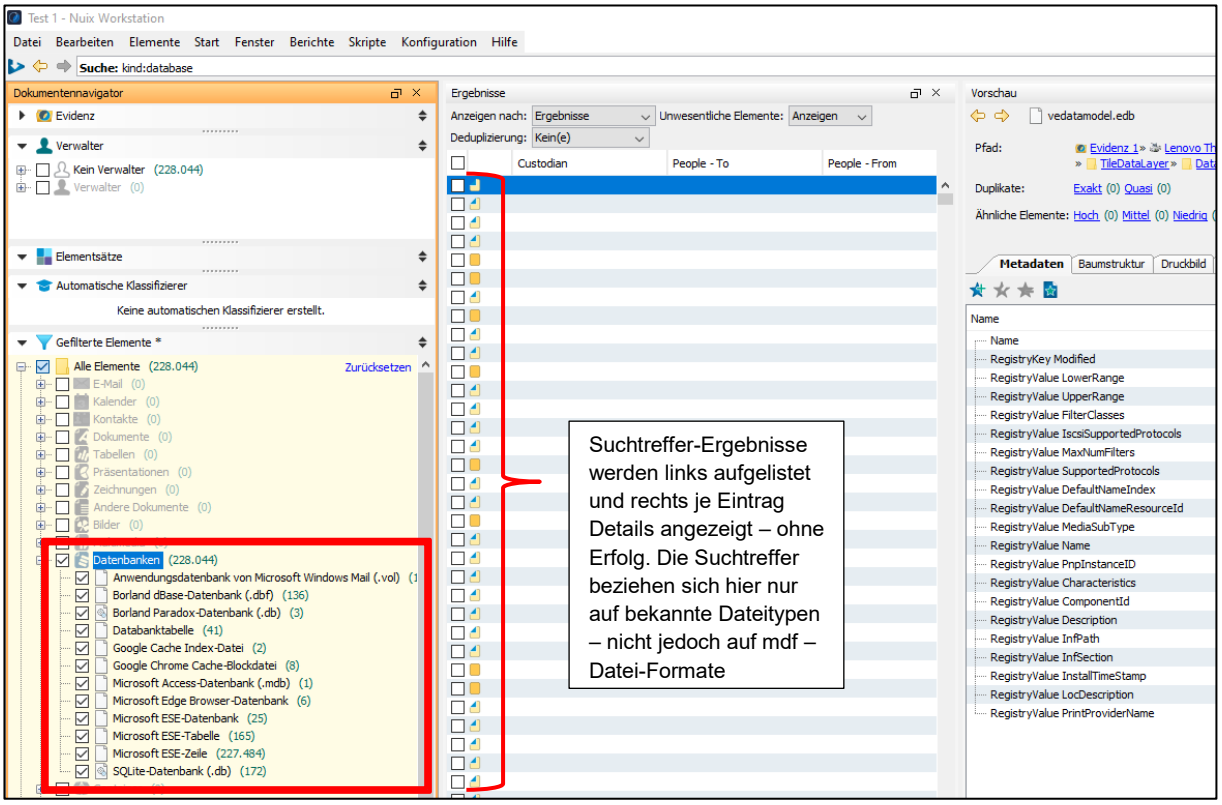


Abbildung 3.31: Verarbeitungsergebnisse für Datenbanktypen unter NUIX WS 8.8

Über verschiedene individuelle Suchbegriffe (Alleinstellungsmerkmale wie Name des Halters, Kennzeichen oder wie in der Abbildung 3.32 die Fahrzeugidentifizierungs-nummer) wurde zusätzlich nach den PDF-Dokumenten im Datenbestand gesucht – ebenfalls ohne Erfolg:

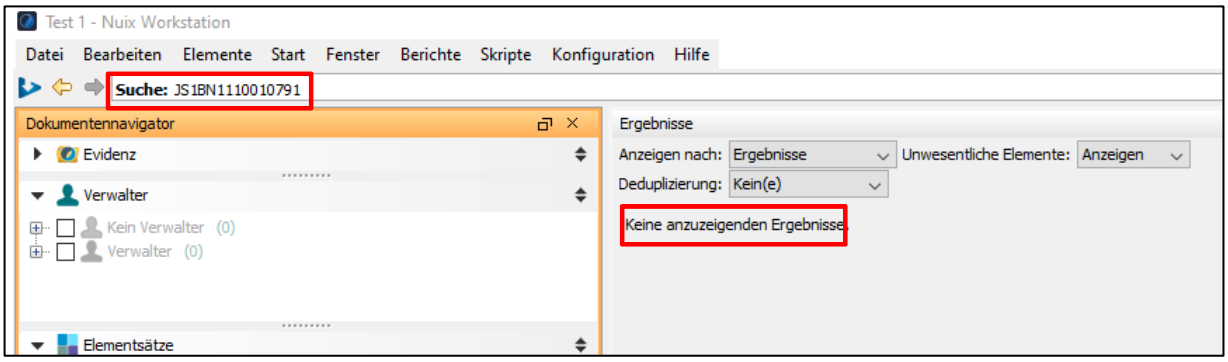


Abbildung 3.32: Suche nach individuellem Begriff unter NUIX WS

Auch NUIX besitzt ebenso wie AXIOM eine Darstellungsansicht der Verzeichnisstruktur und der nachgeordneten Einzeldateien. Im Gegensatz zu AXIOM werden jedoch die nicht erkannten Dateitypen dort nicht mehr angezeigt.

Untersuchung der Protokolldaten unter NUIX Workstation 8.8

Ebenso wie bei AXIOM wird untersucht, ob die Nichtverarbeitung unerkannter File Types gem. der EDRM – Richtlinie *5.5 Ausnahmeberichte* [34] protokollarisch festgehalten wird, um so den Hinweis zu bekommen, dass diese Datentypen gesondert auf Verfahrensrelevanz untersucht werden müssen.

In der Bedienungsanleitung *NUIX Workstation User Guide v8.8.0.pdf* befindet sich auf S. 6 ein Link namens *Supported File Types*, der auf eine Webseite von NUIX verweist (siehe Abbildung 3.33). Doch auch auf dieser Website von NUIX werden keine Anhaltspunkte bzw. Dokumente auf Protokolleinträge für „unsupported Files“ gefunden:

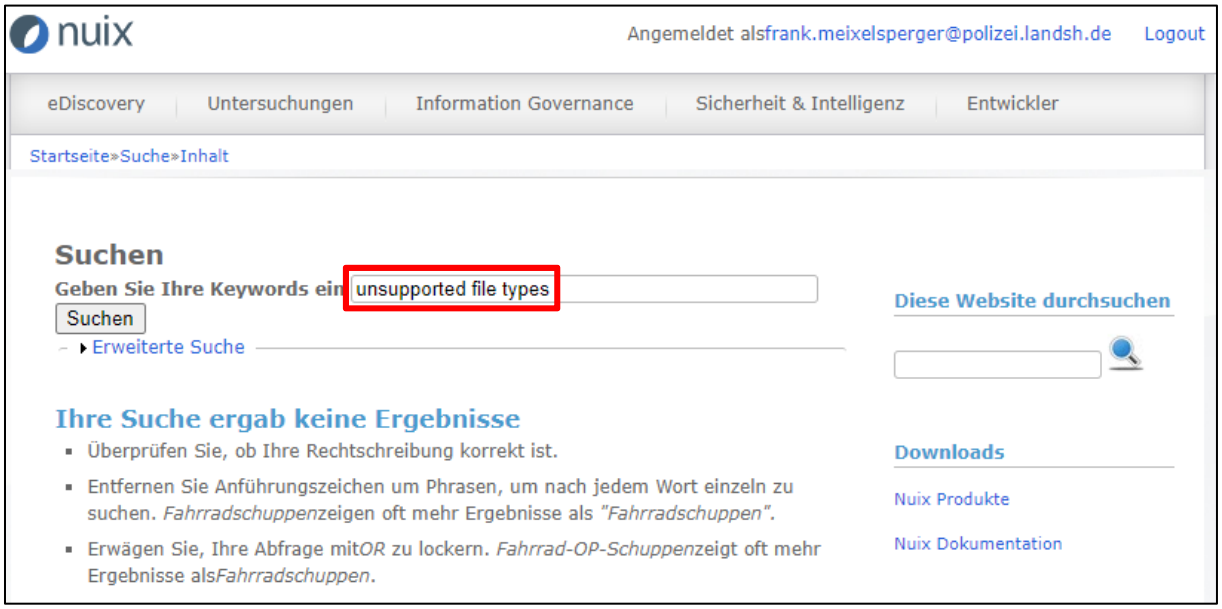


Abbildung 3.33: NUIX-Support Webseite

Auch eine Google-Suche unter NUIX „unsupported file types“ brachte keine relevanten Ergebnisse. Eine Anfrage zu diesem Thema beim deutschsprachigen Support von NUIX wurde bis zur Endfertigung dieser Master Thesis nicht beantwortet.

Abschließend werden die erzeugten NUIX-Protokolldaten zur Verarbeitung des genannten „TÜV“-Images auf Hinweise nicht verabeiteter Dateitypen untersucht.

NUIX protokolliert in vier TXT-Dateien seine Verarbeitungsaktivitäten je Vorgang. Für

[34] https://edrm.net/wiki/5-0-reporting/

jeden eingesetzten Worker (eine einzelne Verarbeitungsengine) werden weitere Protokolle generiert. Für diesen Test standen Lizenzen für 8 Worker zur Verfügung:

Abbildung 3.34: Protokollerstellung in NUIX, hier: je eingesetzten Worker

Eine manuelle Durchsicht all dieser TXT - Protokolldateien brachte die Erkenntnis, dass nicht unterstützte Dateitypen im Rahmen der Verarbeitung ebenfalls nicht protokolliert werden.

Untersuchung der Verarbeitungsergebnisse auf nicht erkannte Dateitypen

NUIX Workstation besitzt zwei MIME-Type Gruppen namens *Unerkannt* und *keine Daten*.

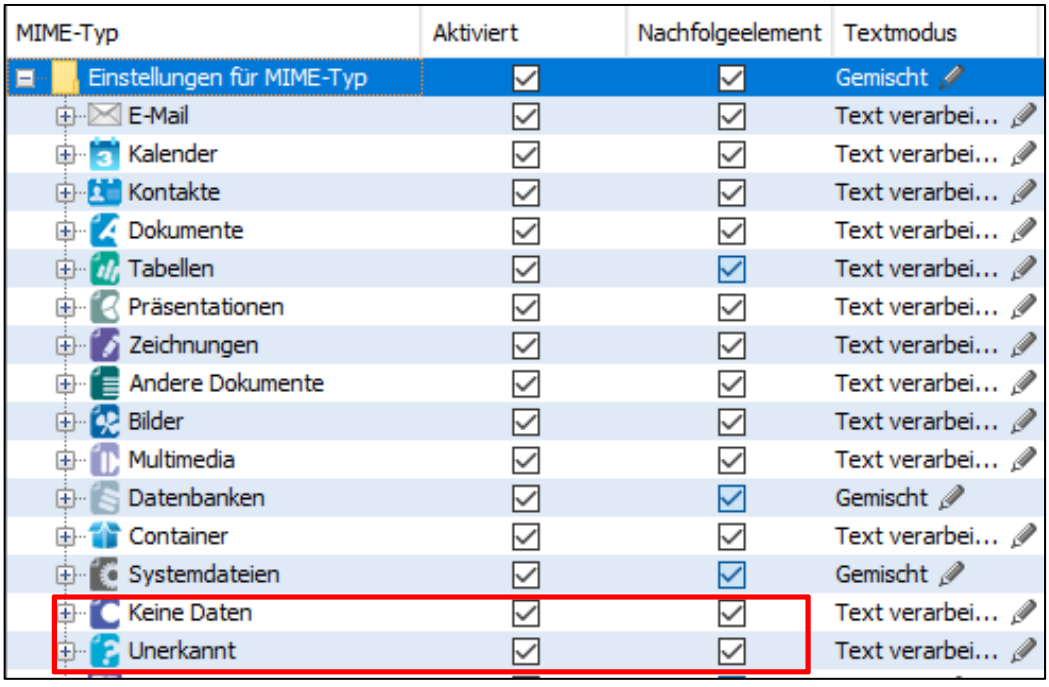


Abbildung 3.35: MIME-Type Einstellungen

Die Verarbeitungsergebnisse zu diesen beiden MIME-Types werden nach der verfahrensrelevanten MDF-Datei durchsucht: Es werden keine Suchtreffer angezeigt (siehe Abbildungen 3.36 und 3.37 nächste Seite).

In den jeweilige Wortlisten (Ergebnis der Indexierung) dieser Suchtreffer wird nach der bereits verwendeten individuellen Zahlenkombination (Fahrzeug-Identifizierungsnr.) gesucht – ebenfalls ohne Erfolg. Dies stützt ebenso die Annahme, dass es für den Dateityp MDF keine Verarbeitungslösung gibt und damit die in dieser Datenbank enthaltenen Filestreams nicht extrahiert und damit auch nicht indexiert werden können.

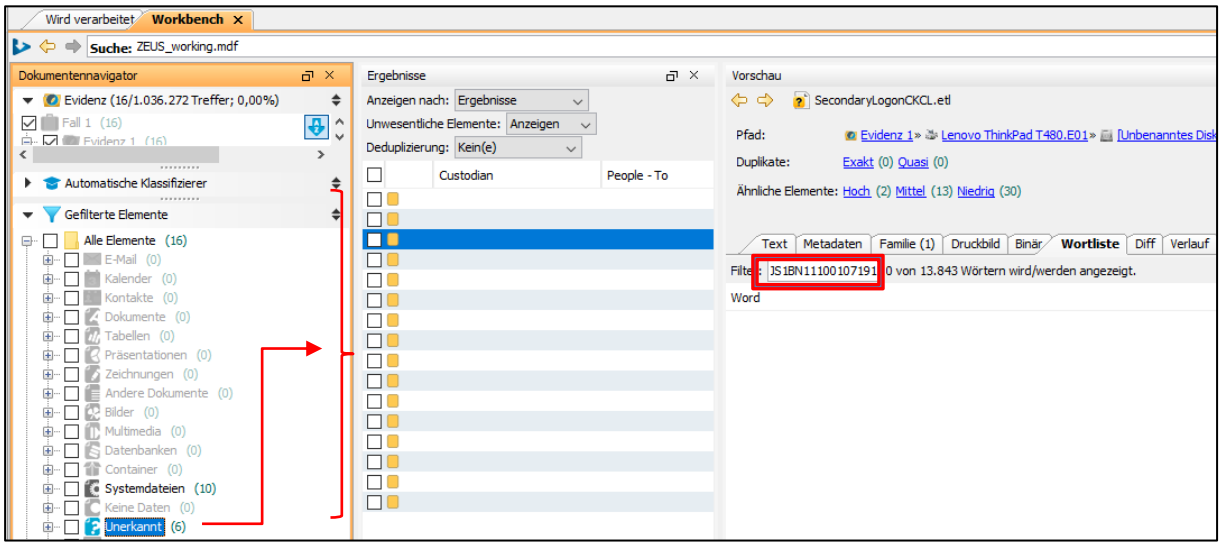


Abbildung 3.36 Suchergebnis für MIME-Type *unerkannt*

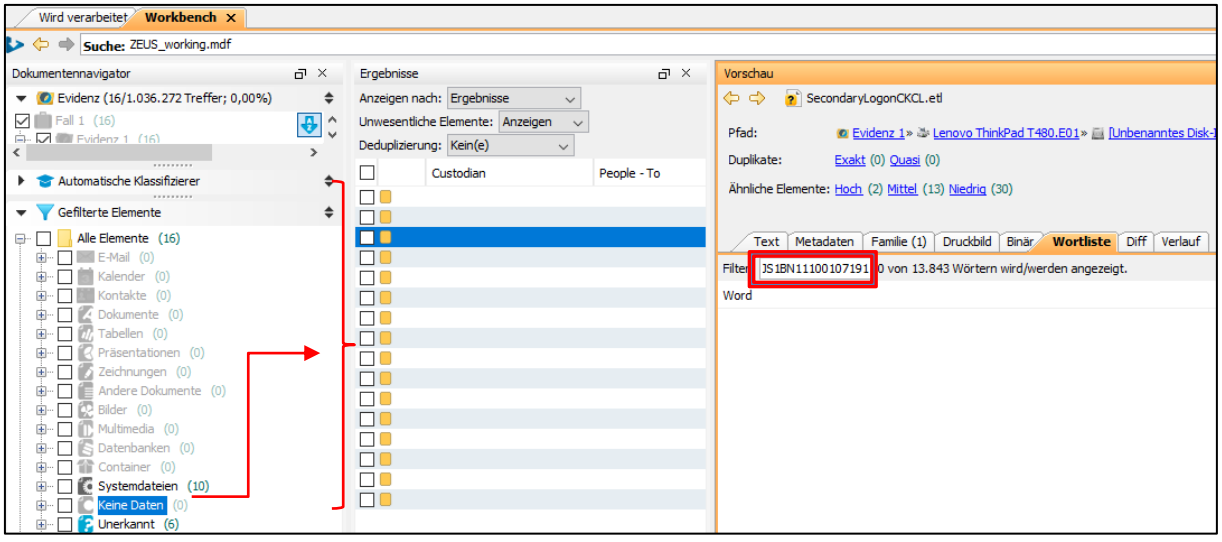


Abbildung 3.37: Suchergebnis für MIME-Type *keine Daten*

Zusammenfassung Test 2

Beide geteste Anwendungen können für nicht bekannte Dateitypen keine Verarbeitungslösung generieren. In den Protokollen beider Anwendungen werden unverarbeitete Dateitypen nicht angemerkt.

Es wird angenommen, dass die festgestellten Mängel ähnlich wie beim Ergebnis zu Test 1 bisher keine Rolle spielten, da eDiscovery-Systeme ursprünglich für US-Zivilprozesse entwickelt wurden und dort dieser Bedarf auf Darstellung nicht verarbeitbarer Datentypen nicht bestand. Dies wird auch durch den Abschnitt 5.5 im Leitfaden 5.0 Berichte [35] deutlich, in dem nicht ausdrücklich die Protokollierung nicht verarbeiteter also unbekannter Dateitypen verlangt wird. Hier wird eine Empfehlung (sollte) ausgesprochen, die auch nicht spezifisch auf nicht erkannte Dateitypen, sondern für alle Gründe der Nichtverarbeitung (Dateibeschädigung, Verschlüsselung, Virusinfektion, NIST-Ausschluss, Kennwortschutz usw.) gilt.

Auch die Prüfung unter NUIX Workstation der MIME-Types *Keine Daten* oder *unerkannt* löst das Defizit nicht, denn entweder sind sie dort nicht vorhanden oder die abgelegten Informationen sind fragmentarisch und geben keinen Hinweis auf ihren Herkunftsort, so dass keine Tatzusammenhänge zum eingesetzten Programm hergestellt werden können.

Bei beiden Tools besteht die Möglichkeit, sich über eine bestimmte Dateiansicht den typischen Verzeichnisaufbau anzeigen zu lassen (siehe hierzu auch Abbildung 3.25). Hier können die installierten Programme oder gespeicherten Daten mit den Verarbeitungsergebnissen im Hinblick auf ihre Nichtverarbeitung verglichen werden. Doch das setzt beim Auswerter ein technisches Verständnis voraus, um diese Defizite einer Nichtverarbeitung erkennen zu können. Zudem wird durch diese “Nach”-Prüfung der Effizienzgrad an Arbeitszeitersparnis gemindert.

Damit bestätigt der Test 2 mit MAGNET AXIOM und NUIX Workstation die Hypothese, dass nicht unterstützte Datentypen, wie in diesem Fall eine SQL-Datenbankdatei vom Typ MDF nicht verarbeitet, also nicht extrahiert, nicht ausgelesen und damit auch nicht indexiert werden.

[35] https://edrm.net/wiki/5-0-reporting/

3.2.4 Test 3 – Verarbeitung von compound files und embedded objects

Vorbemerkung

Im EDRM-Dokument *3.0 Text-, Metadaten- und Bildextraktion* [36] wird in der Einleitung unter anderem folgendes angeführt:

„...*Seriöse Verarbeitungssoftware sollte Feldinformationen, sogenannte Metadaten, aus den Dateien wie Von, An, CC, BCC, Betreff, gesendetes Datum und Uhrzeit sowie Depotbank aus E-Mails extrahieren. Darüber hinaus sollte die Software auch die Möglichkeit bieten, nachverfolgte Änderungen in ein Dokument, ausgeblendete Inhalte und Aktivitäten aufzunehmen. Die extrahierten Informationen können in späteren Phasen des E-Discovery-Prozesses, insbesondere bei der Suche und Überprüfung, von entscheidender Bedeutung sein**...“*

Weiter heißt es im gleichen Dokument unter *3.7 Bilder extrahieren*:

„E-Mail und andere Dateitypen ermöglichen es Inhaltsersteller oft, Bilder und andere Programme in die Datei einzubetten. Fotos und andere Grafiken sind gängige Ergänzungen zu E-Mail- und Office-Dokumenten. Tabellenkalkulationen werden häufig zusätzlich zu Text- oder E-Mail-Nachrichten in Word- oder PowerPoint-Dateien eingebettet.
Verarbeitungssoftware muss in der Lage sein, eingebettete Dateien zu trennen, sie oft ähnlich wie E-Mail-Anhänge zu behandeln und es dem Administrator zu erlauben, zu entscheiden, ob Bilder extrahiert werden oder nicht, da sie ohne Bezug wenig oder gar keinen Wert darstellen...“

Diese Formulierungen verdeutlichen die erhöhten Anforderungen an den Verarbeitungsprozess, da es neben den beschriebenen Dateiextraktionen aus den Schichten 1 - 3 auch die Verzeichnisstruktur von E-Mail-Dateien, also compound files „aufbrechen“ muss, um an die eigentlichen Nachrichten und die Anlagen (Attachments) zu gelangen und sie so für den folgenden Analyseprozess bereitstellen und indexieren zu können. Das gilt ebenso für das Extrahieren von komprimierten Daten und eingebetteten Objekten.
Im Verarbeitungsergebnis (Produkt) sollten die Ergebnisse so dargestellt sein, dass die Abstammung von der Ursprungsdatei oder -dokument eindeutig ist und zugleich auch die in diesen eingebetteten Dateien enthaltenen Meta- und Textinhaltsdaten korrekt wiedergegeben werden.

Der Test 3 geht somit der Frage nach

- ob die in den Originaldaten vorhandenen Meta- u. Inhaltsdaten auch im Endprodukt (Verarbeitungsergebnis) verlustfrei und

[36] https://edrm.net/wiki/3-0text-metadata-and-image-extraction/

- unverändert angezeigt werden oder
- das konvertierte Endergebnis zu tatsächlichen Fehlern oder falschen Interpretationen führt.

Testmaterial 3 – USB 8 GB.E01

Das Testmaterial besteht aus 2 unterschiedlichen Dateien, die nach ihrer Präparierung auf einem USB Stick abgespeichert und dann IT-forensisch gesichert wurden.

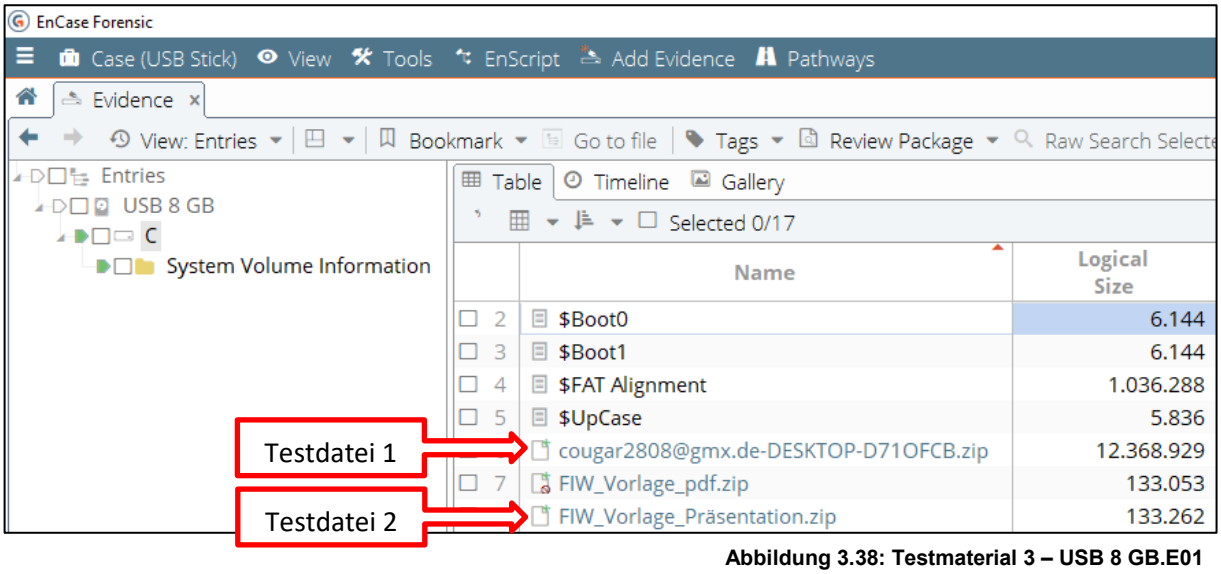


Abbildung 3.38: Testmaterial 3 – USB 8 GB.E01

Dieses Image wird dann für die Tests mit AXIOM und NUIX Workstation eingesetzt.

Testdatei 1 – ZIP-komprimierte MS Outlook-Datei

Durch die zusätzliche ZIP-Komprimierung erschwert, muss die jeweils eingesetzte Verarbeitungsengine die Vielzahl enthaltener Nachrichten, Kontaktdaten, Termine und Anhänge erkennen und korrekt in dem Verarbeitungsergebnis wiedergeben.

Innerhalb dieser PST-Datei befindet sich im Verzeichnisorder *Gelöschte Elemente* eine modifizierte E-Mail mit einer Anlage-Datei namens CP.ZIP. Diese Datei enthält zwei Bilddateien. Wie in Abbildung 3.39 auf der nächsten Seite zu sehen, ist der Dekomprimierversuch selbst mit einem klassischen Forensik-Tool nicht erfolgreich und man kann die Bilder nicht einsehen. Nichts weist daraufhin, dass diese Datei verschlüsselt und mit einem Passwort versehen ist.

Aufgrund der Tatsache, dass diese Nachricht im IT Forensik-Tool Encase nach der Analyse und Datenaufbereitung im Ordner *Gelöschte Elemente* liegt, kann der Eindruck entstehen, dass dies Ursache der Fehldarstellung ist. Durch den Löschprozess könnte der Inhalt dieser Mail und sein Attachement vollständig oder teilweise über-

schrieben und damit nicht mehr einsehbar sein:

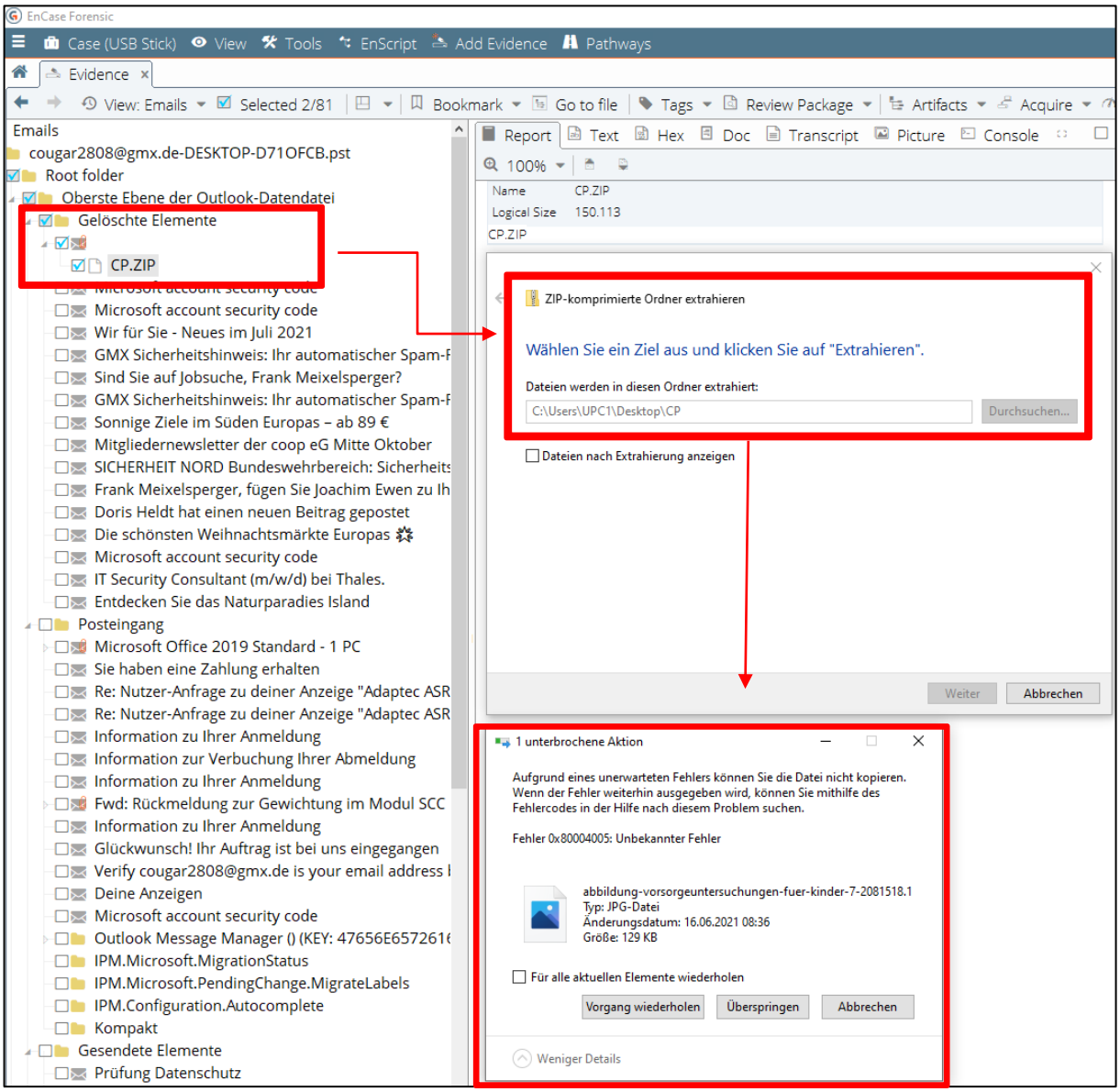


Abbildung 3.39: Datenextraktion einer PST-Datei unter Encase Forensics 8

Nur durch die Analyse des Headers dieser ZIP – Datei würde erkennbar werden, dass es sich um eine passwortgeschützte Datei handelt:

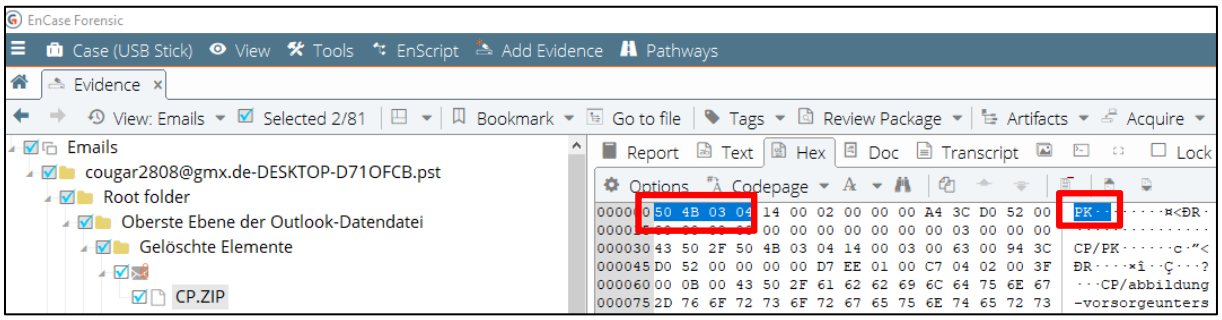


Abbildung 3.40: Matching Number für das PKZIP Format

Eine Abfrage der „matching number“ 50 4B 03 04 unter der Website https://asecurity-site .com/forensics/magic führt zu folgendem Ergebnis:

Windows Audio file	.wma	30 26 B2 75 8E 66 CF
PKZip	.zip	50 4B 03 04 [PK]
GZip	.gz	1F 8B 08

Abbildung 3.41: Header-Signatur von CP.ZIP

Damit wird deutlich, es handelt sich hier nicht nur um eine Komprimierung, wie die Dateiendung *.ZIP suggeriert, sondern zusätzlich um eine passwortgeschützte und verschlüsselte ZIP-Datei. Mit diesem Wissen kann man nun versuchen, die Verschlüsselung über einen Bruteforce - Angriff zu decryptieren und die 2 enthaltenen Bilder sichtbar zu machen:

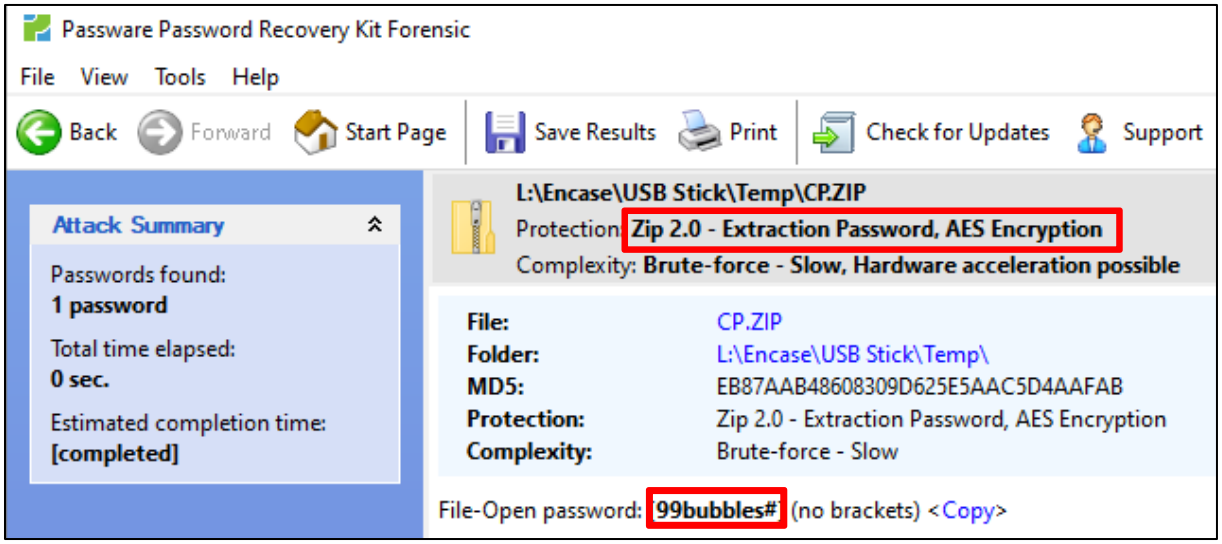


Abbildung 3.42: Erfolgreicher Bruteforce-Angriff auf CP.ZIP

Selbst für einen erfahrenen IT-Forensiker kann schnell mit einem klassischen IT-Forensik-Tool wie Encase diese Verschlüsselung übersehen werden, wenn man die Ursache der Fehlfunktion nicht näher durch die Analyse der Headerinformation verifiziert.

Abbildung 3.43: Dekomprimierung und Decryptierung einer Bilddatei aus dem Ordner CP.ZIP

Wie gut oder schlecht eDiscovery-Anwendungen mit diesen Besonderheiten umgehen, soll der Test zeigen, denn laut EDRM-Dokument *1.0 ESI-Aufnahme und Dateiextrak-tion* [37] sollen eDiscovery-Anwendungen während des Analyse- und Erzeugungsprozesses diese Funktion unterstützten – eine Header-Analyse zur Identifizierung von Dateitypen.

Testdatei 2 – Komprimierte Präsentationsdatei mit eingebettetem Objekt

Präparierung einer Powerpoint-Datei für den Test durch:

a) Verwendung einer WINGS- Mustervorlage

Abbildung 3.44: Powerpoint-Datei

b) Erzeugung einer Excel-Datei namens *Test1 eDiscovery.xlsx* mit 2 Tabellen in den Kartenreitern *Tabelle1* und *Tabelle 2*

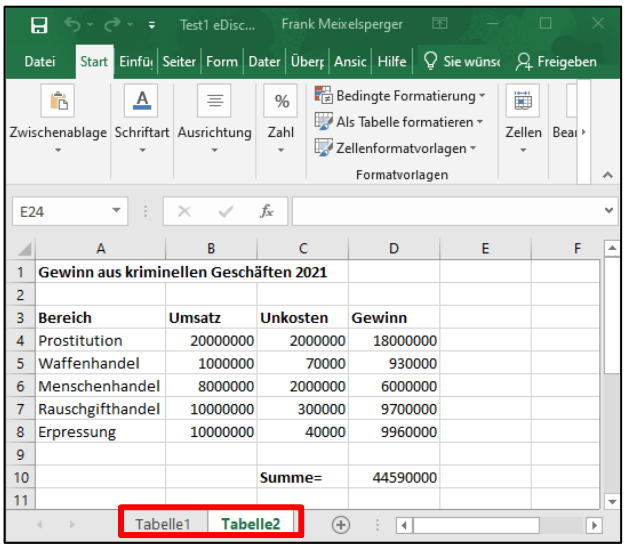


Abbildung 3.45: Excel-Datei mit 2 Tabellen

c) Einbettung der Excel-Tabelle *Test1 eDiscovery.xlsx* mit 2 Tabellen über den Menübefehl *Format /Objekt* in die Präsentation.

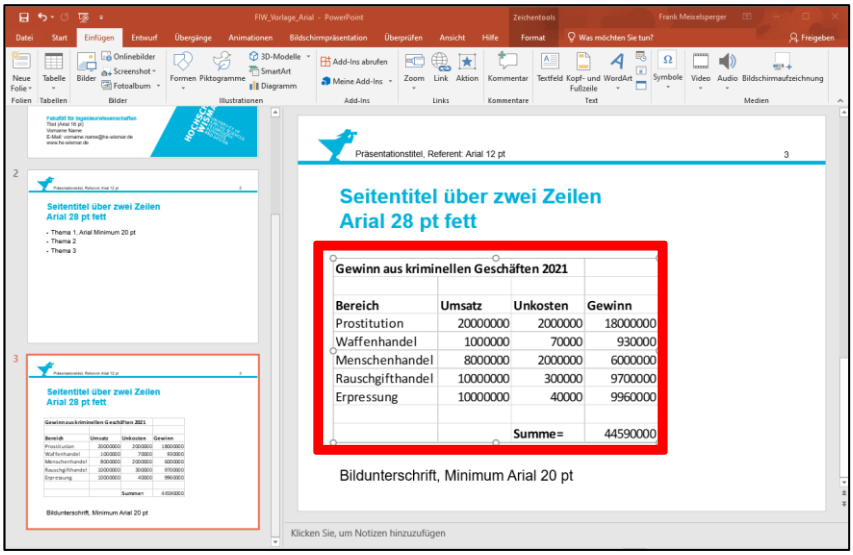


Abbildung 3.46: Powerpoint-Datei mit eingebetteter Tabelle

[37] https://edrm.net/wiki/1-0-esi-ingestion-and-file-extraction/

Das Ziel dieses Tests ist es herauszufinden, ob die embedded Objects in der Präsentationsdatei, also die beiden Excel-Tabellen, im Verarbeitungsergebnis beider Testtools korrekt extrahiert und analysiert wieder gegeben werden. Als Referenz dient das Prüfsummenverfahren und die Zeitstempelanalyse. Eine fehlerhafte Extraktion würde falsche Informationen hierzu liefern.

Untersuchungsergebnisse mit AXIOM 6.9

Testdatei 1: cougar2808@gmx.de_DESKTOP-D710FCB.pst

Zwar wurde die E-Mail-Datei durch AXIOM in ihrer Gesamtheit nicht „gehasht“, dafür funktionierte die Prüfsummenbildung für alle enthaltenen E-Mails trotz der „compound“-Schwierigkeit aus ZIP- und PST-interner Komprimierung einwandfrei. Ein stichpunktartiger Vergleich von 3 Prüfsummen generiert mit Encase Forensics war korrekt, es gab keine Differenzen in den Hash-Werten:

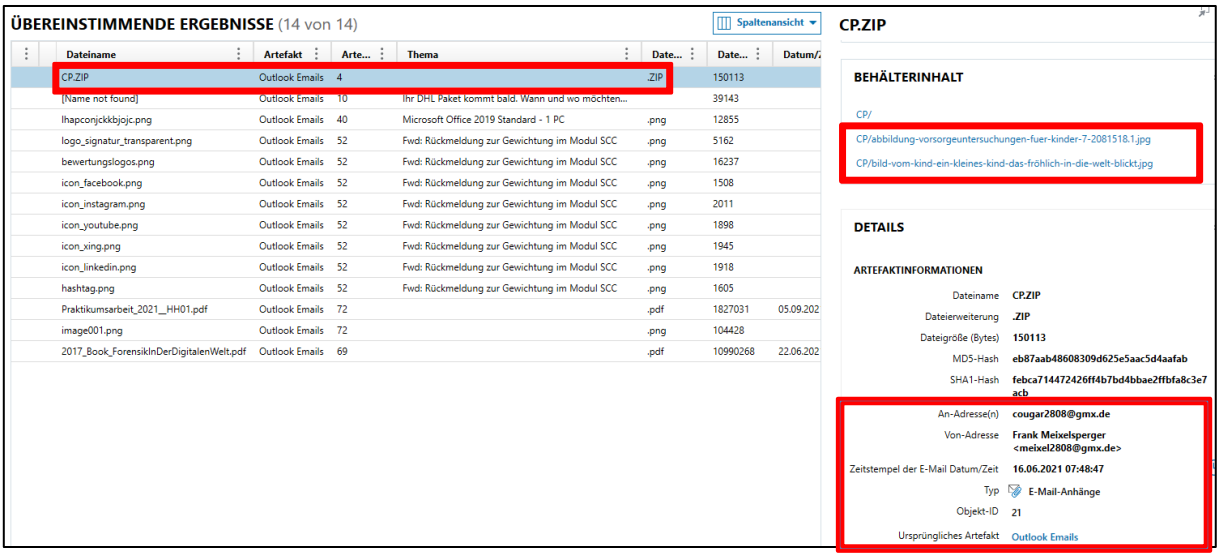


Abbildung 3.47: Verarbeitungsergebnis der Datei CP.ZIP mit AXIOM 6.8

Die genauere Untersuchung des Verarbeitungsergebnisses zur speziellen Datei CP.ZIP zeigte auch korrekte Hash- und Timestamp-Werte an, so dass diese Informationen sicher zur Suche in anderen Datensicherungen eingesetzt werden könnten. Auch wurden die Inhaltsinformationen - zwei JPG Bilddateinamen (siehe rechte obere rote Markierung in Abbildung 3.47) - korrekt angezeigt, konnten jedoch nicht geöffnet bzw. eingesehen werden. Weder gab es eine Editorfunktion wie bei Encase, mit der man hätte den Header der ZIP-Datei einsehen können, noch wurde auf eine Verschlüsselung über die erwähnten Protokolle hingewiesen.

Der Versuch diese zwei Bilder für weitere Untersuchungen zu exportieren und per Dritt-Tools zu untersuchen, mißlang ebenfalls wegen der bekannten Verschlüsselung:

Abbildung 3.48: Fehlermeldung in Axiom 6.5

Testdatei 2: FIW_Vorlage_pptx_.pptx

Die in die Präsentationsdatei eingebettete Excel-Datei wird mit beiden Tabellen und auch korrekten Metadaten der Excel-Tabelle angezeigt:

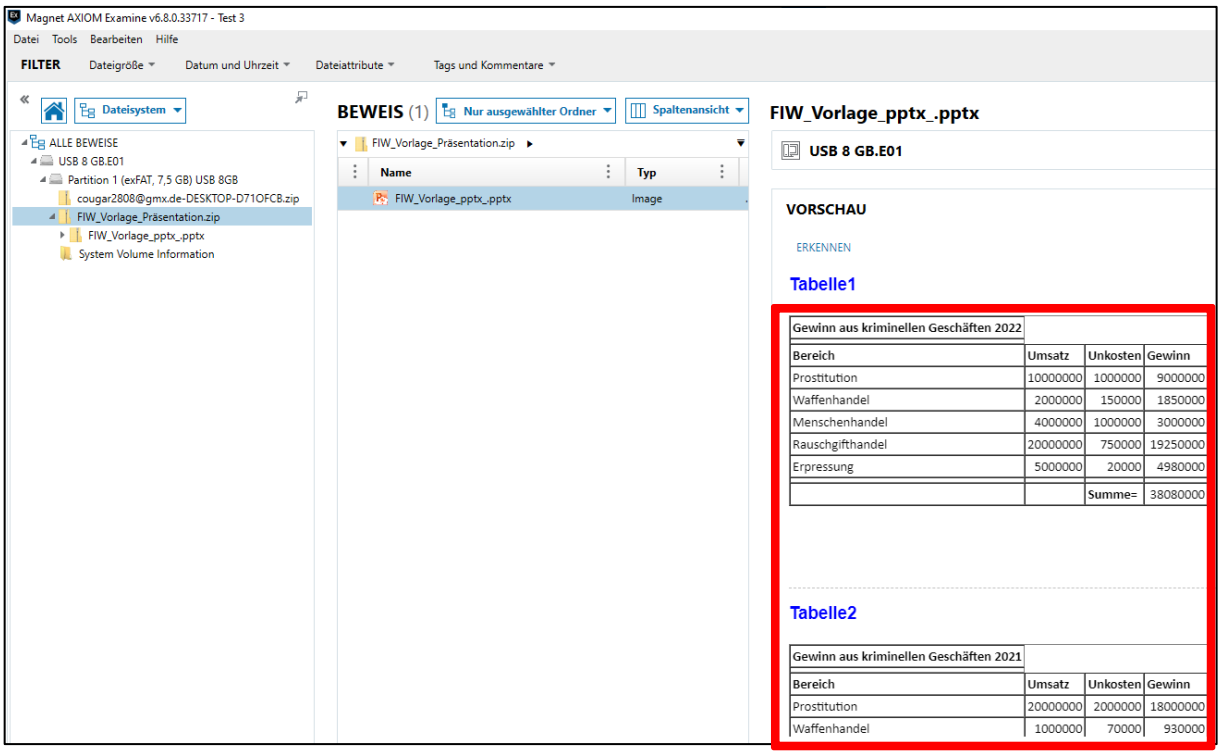


Abbildung 3.49: Verarbeitungsergebnis unter AXIOM Examine

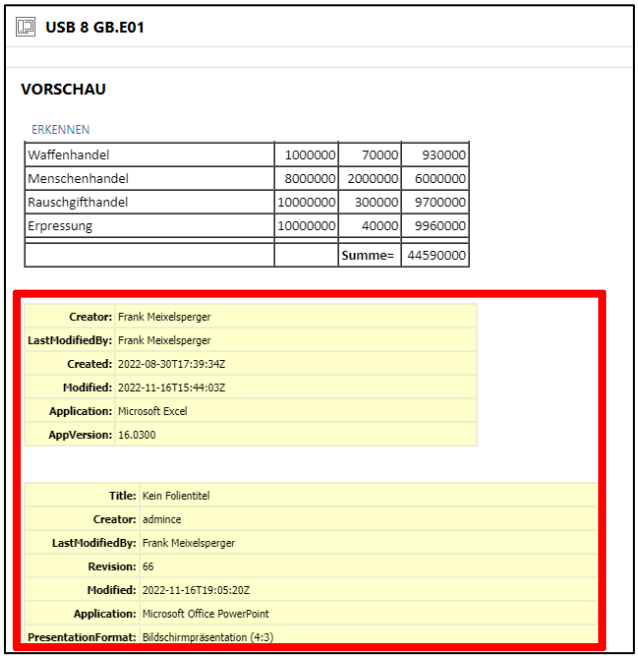


Abbildung 3.50 Metadaten-Anzeige des „embedded object“

Auch der Vergleich der Hashwerte der „gezippten“ Powerpoint-Datei und der wiederum eingebetten Excel-Datei ergab sowohl in Encase als auch AXIOM die gleichen Werte:

AXIOM	Encase
ARTEFAKTINFORMATIONEN Dateiname: **FIW_Vorlage_pptx_.pptx** Datum/Zeit der letzten Modifikation auf das Dateisystem: **16.11.2022 21:05:20** Größe (Bytes): **151224** Titel: **Kein Folientitel** Gespeicherte Größe (Bytes): **151224** Autoren: **admince** Letzter Autor: **Frank Meixelsperger** Zuletzt modifiziert – Datum/Zeit: **16.11.2022 20:05:20** MD5-Hash: **a895272efc70dec09afd9dba5cfb71e0** SHA1-Hash: **d416948d103bb6f01f52ede2df5a85b26fc9e95f** Typ: PowerPoint-Dokumente	Name: FIW_Vorlage_pptx_.pptx File Ext: pptx Logical Size: 151.224 Category: Document - Presentation Signature Analysis: Match File Type: Microsoft Powerpoint Template 2007-2010 Last Written: 16/11/22 20:05:20 (+1:00 Mitteleuropäische Zeit) MD5: a895272efc70dec09afd9dba5cfb71e0 SHA1: d416948d103bb6f01f52ede2df5a85b26fc9e95f Item Path: FIW_Vorlage_pptx_.pptx True Path: USB Stick\USB 8 GB\C\FIW_Vorlage_Präsentation.zip\FIW_Vorlage_pptx_.pptx
ARTEFAKTINFORMATIONEN Dateiname: **Microsoft_Excel_Worksheet.xlsx** Datum/Zeit der letzten Modifikation auf das Dateisystem: **01.01.1980 01:00:00** Größe (Bytes): **13000** Gespeicherte Größe (Bytes): **13000** Autoren: **Frank Meixelsperger** Letzter Autor: **Frank Meixelsperger** Zuletzt modifiziert – Datum/Zeit: **16.11.2022 16:44:03** Datum/Zeit der Erstellung: **30.08.2022 19:39:34** MD5-Hash: **7a2dbca92b827e55a4458167c6f2cd64** SHA1-Hash: **e202a52870da4c7b448111840de39bb725063e54** Typ: Excel-Dokumente	Name: Microsoft_Excel_Worksheet.xlsx File Ext: xlsx Logical Size: 13.000 Category: Document - Spreadsheet Signature Analysis: Match File Type: Microsoft Excel Spreadsheet 2007-2010 Last Written: 01/01/80 00:00:00 (+1:00 Mitteleuropäische Zeit) MD5: 7a2dbca92b827e55a4458167c6f2cd64 SHA1: e202a52870da4c7b448111840de39bb725063e54 Item Path: ppt\embeddings\Microsoft_Excel_Worksheet.xlsx True Path: USB Stick\USB 8 GB\C\FIW_Vorlage_Präsentation.zip\FIW_Vorlage_pptx_.pptx\ppt\embeddings\Microsoft_Excel_Worksheet.xlsx

Tabelle 2: Vergleich der Hashwerte von Encase und Axiom

Untersuchungsergebnisse mit NUIX Workstation 8.8

Testdatei 1: cougar2808@gmx.de_DESKTOP-D710FCB.pst

Auch in dieser eDiscovery-Anwendung lässt sich die komprimierte Outlook – Datei nicht hashen. Es werden nur für die enthaltenen E-Mails MD5-Hashwerte generiert. Die passwortgeschützte Datei CP.ZIP befindet sich im Verarbeitungsergebnis:

Abbildung 3.51: Verarbeitungsergebnis des E-Mail Attachements in NUIX WS 8.8

Und auch die komprimierten Bilddateinamen werden korrekt angezeigt, können aber nicht geöffnet werden.

Im Gegensatz zu AXIOM erfolgt hier aber ein Hinweis auf eine Verschlüsselung:

Abbildung 3.52: Hinweis auf eine verschlüsselte Datei unter NUIX

Durch eine andere Darstellungsansicht (hexadezimal) lässt sich eine Header-Analyse durchführen und damit der Hinweis auf einen Passwortschutz (PKZIP) erlangt werden (siehe nächste Abbildung 3.52):

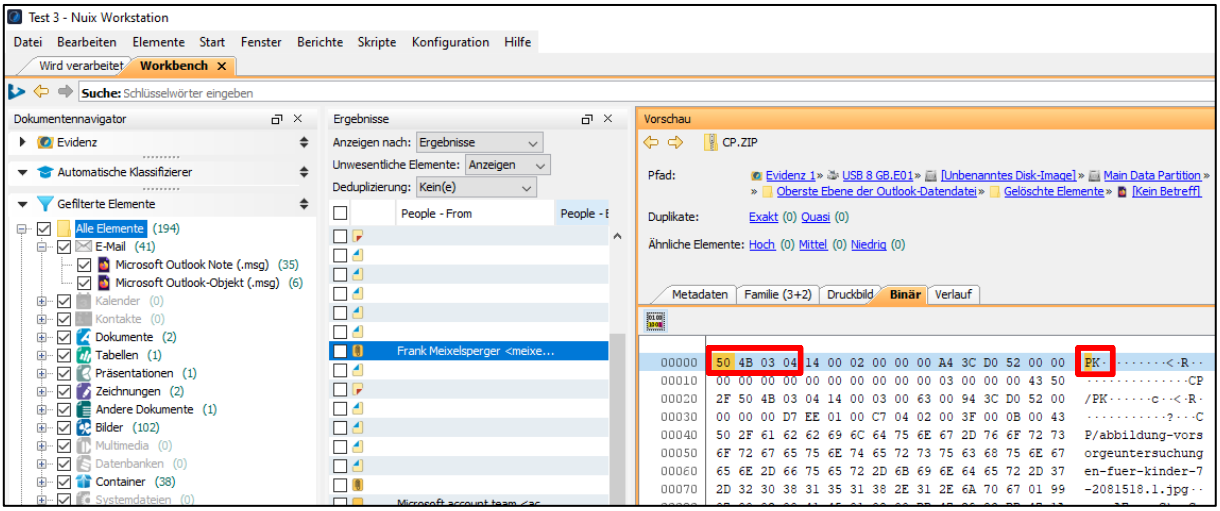


Abbildung 3.53: HEX-Ansicht unter NUIX

Testdatei 2: FIW_Vorlage_pptx_.pptx

Die in die Powerpoint-Präsentation eingebettete Excel-Datei wird im Verarbeitungsergebnis angezeigt, jedoch als einzelnes Objekt und nicht wie bei AXIOM als Teil der Powerpoint-Datei. Es erfolgt allerdings ein Hinweis, dass diese Excel Worksheet in die Präsentationsdatei eingebettet ist (siehe rote Markierung, Abbildung 3.54):

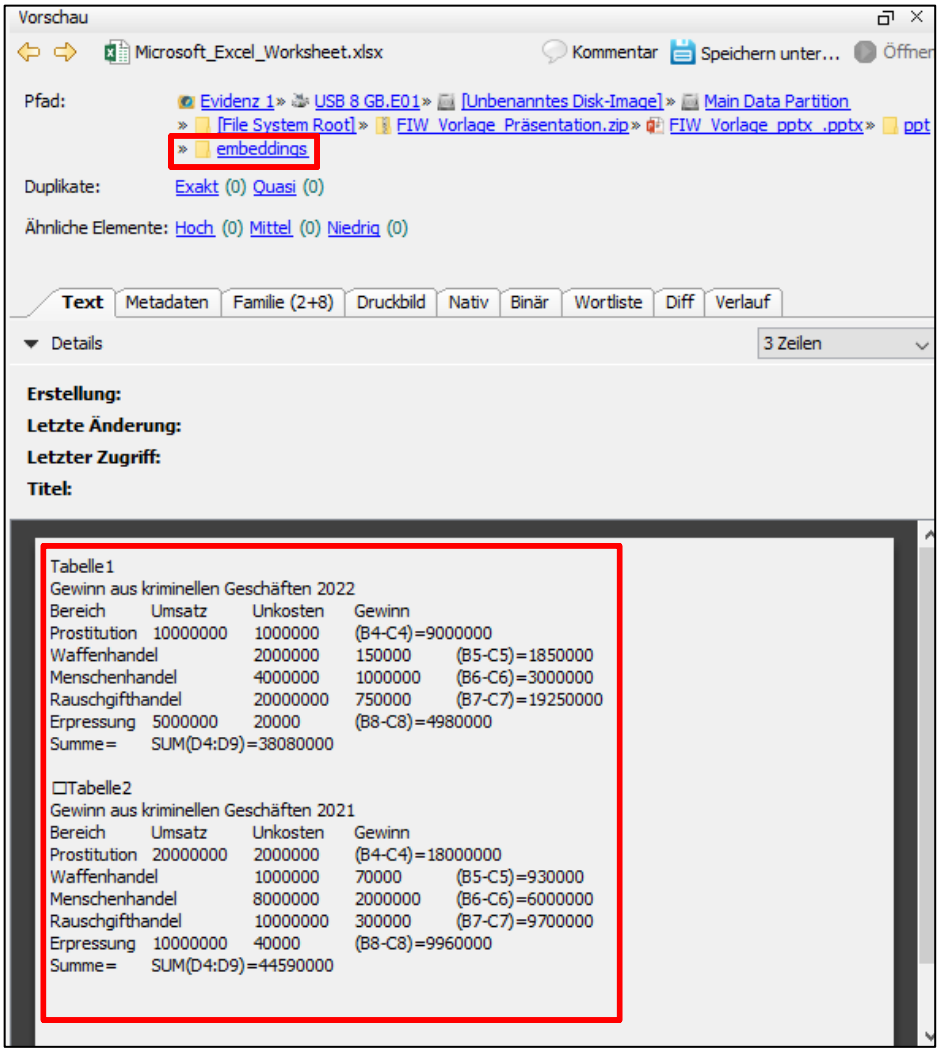


Abbildung 3.54 Darstellung von eingebetteten Objekten unter NUIX

Die MD5-Hashwerte sowohl der Powerpoint-, als auch der eingebetteten Excel-Datei sind korrekt wie unter Encase oder AXIOM berechnet. Teilweise werden sogar mehr Timestamp-Informationen als unter Encase ausgelesen, wie die nachfolgende Gegenüberstellung zeigt:

NUIX	Encase
Name FIW_Vorlage_pptx_.pptx Last Saved Mittwoch, 16. November 2022 um 20:05:20 Mitteleurop File Modified Mittwoch, 16. November 2022 um 20:05:20 Mitteleur MD5-Übersicht a895272efc70dec09afd9dba5cfb71e0	Name FIW_Vorlage_pptx_.pptx File Ext pptx Logical Size 151.224 Category Document - Presentation Signature Analysis Match File Type Microsoft Powerpoint Template 2007-2010 Last Written 16/11/22 20:05:20 (+1:00 Mitteleuropäische Zeit) MD5 a895272efc70dec09afd9dba5cfb71e0 SHA1 d416948d103bb6f01f52ede2df5a85b26fc9e95f Item Path FIW_Vorlage_pptx_.pptx True Path USB Stick\USB 8 GB\C\FIW_Vorlage_Präsentation.zip\FIW_Vorlage_pptx_
Name Microsoft_Excel_Worksheet.xlsx Last Saved Mittwoch. 16. November 2022 um 16:44:03 Mi Created Dienstag, 30. August 2022 um 19:39:34 Mitteleuropäisc File Modified Mittwoch, 16. November 2022 um 20:05:20 Mitt MD5-Übersicht 7a2dbca92b827e55a4458167c6f2cd64	Name Microsoft_Excel_Worksheet.xlsx File Ext xlsx Logical Size 13.000 Category Document - Spreadsheet Signature Analysis Match File Type Microsoft Excel Spreadsheet 2007-2010 Last Written 01/01/80 00:00:00 (+1:00 Mitteleuropäische Zeit) MD5 7a2dbca92b827e55a4458167c6f2cd64 SHA1 e202a52870da4c7b448111840de39bb725063e54 Item Path ppt\embeddings\Microsoft_Excel_Worksheet.xlsx True Path USB Stick\USB 8 GB\C\FIW_Vorlage_Präsentation.zip\FIW_Vorlage_pptx_.pptx\ppt\embeddings\Mic _Excel_Worksheet.xlsx

Tabelle 3: Gegenüberstellung der Metadaten aus NUIX und Encase

Zusammenfassung Test 3

Die in den Originaldaten vorhandenen Meta- und Inhaltsdaten werden, soweit es hierfür eine Verarbeitungslösung gibt (erkannte Dateitypen), auch im Endprodukt (Verarbeitungsergebnis) verlustfrei dargestellt. Das gilt insbesondere für Hashwerte und Zeitstempel. Auch durch die verschiedenen Komprimierungen ließen sich beide Tools nicht „irritieren“ und zeigten korrekte Werte wie der Vergleich mit der Referenz Encase zeigt an.

Beide Tools können das embedded Object einer Excel-Datei in einer Powerpoint-Präsentation im Verarbeitungsergebnis darstellen. Auch die enthaltenen Textinformationen wurden korrekt indexiert.

NUIX wies auf eine Verschlüsselung für die Datei CP.ZIP hin, AXIOM nicht. Die Darstellungsweise des Verarbeitungsergebnisses wird dafür in AXIOM deutlich übersichtlicher gestaltet.

4 Zusammenfassung

4.1 Nutzbarkeit von eDiscovery-Systemen für Strafverfahren

4.1.1 Veränderung des Originalmaterials

Die Tests mit den beiden eDiscovery-Programmen haben ergeben, dass die Konvertierung der Originaldaten in ein neues Verarbeitungsergebnis bzw. eine Datenbank hinsichtlich ihres Informationsgehaltes (Eigentümer/Nutzer der Datei, Herkunftsort der Datei, Zeitstempel etc.) nicht verändert werden.

4.1.2 Vollständigkeit der Untersuchung

Die in den Grundlagen erklärte Verarbeitungsmethode von eDiscovery-Systemen führte zu der Annahme, dass nicht alle Daten aus logischen und forensischen Sicherungen von einem derartigen System verabeitet und damit für die Auswertung und Recherche sichtbar gemacht bzw. zur Verfügung gestellt werden können.
Mit den Untersuchungsergebnissen aus dem Test 2 konnte diese Hypothese belegt werden.

Erhöht wird dieses Risiko durch den vom EDRM Gremium empfohlenen Einsatz von Culling – Maßnahmen vor dem eigentlichen Processing, der zur Reduzierung der Verarbeitungsmenge und damit Verarbeitungszeit dient. Während das Wegfiltern von Dubletten (je nach Untersuchungsziel) noch als unkritisch betrachtet wird, ist der Einsatz von DeNISTing vor der eigentlichen Verarbeitung kritisch zu sehen. Denn diese Maßnahme entzieht dem Auswerter die Möglichkeit, über z.B. Zeitstempel-Analysen die Nutzung von Programm- und Systemdateien festzustellen und damit ihren Kontext zum Nutzungsablauf auf dem Quellgerät herzustellen. Aktivitäten und Handlungsabläufe im Zusammenhang mit dem Tatgeschehen bzw. Tathandlungen sind dann nur eingeschränkt möglich.

Durch diesen im EDRM-Gremium allgemein beschriebenen Workflow der aus Sicht des Autors ungeprüften Datenreduktion und eigenständigen vollautomatisierten Filterungen und Aufbereitungen von Daten wird die Entscheidungskompetenz, was Verfahrensrelevanz besitzt, an die programmierte eDiscovery-Lösung „abgegeben“. Diese kann aber nur innerhalb ihrer programmierten Parameter Lösungen anbieten. Es bleibt somit eine Differenz an Datenmenge aus den Datensicherungen, die einer

Aus- und Bewertung durch die Auswerter/Ermittler entzogen wird.
Als weiteres Risiko bei der späteren Auswertung wird die vollautomatische aber unvollständige Datenaufbereitung angesehen. Zwar haben Tools wie NUIX einen guten Umfang von über 1000 unterstützen Dateitypen, doch ein Restrisiko bleibt. Eine eindeutige und übersichtliche Protokollierung nicht verarbeiteter Dateien bzw. -typen würde hier Abhilfe schaffen.
Doch diese findet in beiden geprüften Tools nicht oder nur fragmentarisch und nicht eindeutig in ihrer Darstellung statt. Die Leitfäden des EDRM-Gremiums schreiben diese Ausweisung unerkannnter Dateitypen in Protokollen nicht explizit vor. Stattdessen werden als Begründung für Protokollierungen von fehlerhaften Verarbeitungen z.B. der Virenbefall oder defekte Dateitypen zitiert, nicht jedoch unbekannte Dateitypen.

Aufgrund der beruflichen Praxis des Autors und seine Erfahrung in der Zusammenarbeit mit Ermittlern / Auswertern wird das Risiko als groß angesehen, dass unverarbeitete Datentypen durch diese Personengruppe nicht erkannt werden und die dafür notwendige zusätzliche klassische IT-Forensik-Untersuchung deshalb auch nicht beauftragt wird.

4.1.3 Effizientere Auswertung bei großen Datenmengen und Mehrwerte

Durch die Fähigkeit, große Datenmengen zu verarbeiten und vorzuhalten, ergeben sich gerade bei komplexen Verfahren mit vielen Tatbeteiligten Mehrwerte und neue Lösungsansätze, die mit klassischen IT-Forensik-Bearbeitungsformen nicht oder nur unter sehr großem Personaleinsatz möglich wären. Die Harmonisierung von Dateninhalten und Umkonvertierung der Textinhalte in eine oder mehrere Indexlösungen erleichtert erheblich das gezielte Suchen nach verfahrensrelevanten Inhalten und stößt das Tor zur Nutzung von *Big Data Analysis*-Methoden und dem Einsatz von *Künstliche Intelligenz (KI)*-Techniken auf.
Damit können z.B. Kommunikationsbeziehungen zwischen Tätern sichtbar gemacht werden, wenn in unterschiedlichen forensischen Sicherungen, die in einer eDiscovery-Lösung zusammengeführt sind, gleiche Kommunikationsdaten (z.B. Nicknames in Instant Messengern oder Kontaktdaten in Telefonbüchern) gefunden werden.

Big Data-Analysen dienen in der Wirtschaft vor allem dazu, Trends vorherzusagen, wohin sich z.B. ein Bedarf entwickelt, indem man innerhalb der Gesamtenge vorhan-

dener Daten nach bestimmten gleichen Werten bzw. Parametern von Nutzern sucht und durch Gegenüberstellung und Vergleich eine Entwicklung, eine neue „Bedarfsrichtung der Kunden“ erkennen kann. Diese Technik, gleichartige Dateninhalte (Telefonkontakte, Nicknames bei Instant Messengern, Geokoordinaten, E-Mail-Adressen usw.) von unterschiedlichen Spurenlegern/Tätern zu vergleichen, können für Analysen bei polizeilichen Ermittlungen genutzt werden. Sie helfen

- über die Einsammlung aller zu einem Tatbeteiligten vorhandenen Daten aus der Gesamtmenge,
- die Kommunikationsbeziehungen von Tätern untereinander zu visualisieren oder
- Täterverhalten und die Rolle eines Täters innerhalb einer kriminellen Vereinigung zu erkennen und/oder besser zu verstehen

und damit

- schneller als mit herkömmlichen Methoden

wichtige Erkenntnisse für die Entscheidung von Folgemaßnahmen oder zur Aufklärung des Sachverhaltes zu erlangen.

4.1.4 Vergleich

Gegenüberstellung der Vor- und Nachteile beider Modelle:

SAP-Modell	
Vorteile	**Nachteile**
• Genauigkeit der Untersuchung • Vollumfänglichkeit der Untersuchung • Hohe Wahrscheinlichkeit, dass verfahrensrelevante Spuren gefunden werden • Gerichtsfest, anerkannte Arbeitsabläufe mit teilweise evaluierten Forensik-Tools	• hoher Ressourcenverbrauch bei Personal- u. Finanzmitteln für Anschaffung von Hard- u. Software sowie Planung und Aufbau einer notwendigen IT-Infrastruktur • Lange Wartezeiten zur Datenaufbereitung bzw. Ergebniserstellung bei großen Datenmengen und unbekannter neuer Software • Big Data-Analysen mit klassischen Analysen nicht möglich, da nicht geeignet für die Verwaltung großer Datenmengen

eDiscovery-Modell	
Vorteile	**Nachteile**
• Hohe Effizienz bei der Auswertung großer Dokumentenbestände und damit Einsparpotential bei Arbeitsaufwänden, schnellere Erkenntnisgewinnung • Bei Nutzung in einer Netzwerklösung sehr effizienter Einsatz von Personal durch parallele Auswertung • Stabil und schnell in der Vorhaltung großer Datenmengen • Durch Automation der Datenaufbereitung Einsparpotential bei Personalressourcen • Mehrwerte durch Big Data-Analysen und damit zusätzliche Erkenntnisgewinnungen	• Keine vollumfängliche Untersuchung möglich, limitiert durch den Leistungsumfang der jeweils eingesetzten eDiscovery-Anwendung, Risiko, dass verfahrensrelevante Spuren nicht gefunden werden • Nicht geeignet für alle Deliktsbereiche • Akzeptanz des Untersuchungsverfahrens bei deutschen Gericht noch unklar, bisher keine Evaluation durch Behörden erfolgt • Hoher Finanzmitteleinsatz für Anschaffung und Betrieb von Hard- und Software • Lange Verarbeitungszeiten • Zusätzliches Fachpersonal für Analysen und Skriptprogrammierung notwendig

Tabelle 4: Vor- und Nachteile im SAP- und eDiscovery-Modell

4.2 Empfehlungen und Ausblick

4.2.1 Empfehlungen

Die Frage dieser Master-Thesis, ob eDiscovery – Systeme für deutsche Strafverfahren einsetzbar sind, konnte beantwortet werden. Ja - aber nicht uneingeschränkt!

Bei der Auseinandersetzung mit dieser Technologie wurde aber auch deutlich, dass die Anbieter dieser Software stetig den Leistungsumfang ihrer eDiscovery-Produkte ausbauen und damit das Risiko zukünftig kleiner wird, verfahrensrelevante Spuren nicht aufbereitet zu bekommen, wodurch sich die Qualität verbessert und die Einsatzmöglichkeiten für diese Software-Systeme erweitert bzw. der Umfang verarbeitbarer Datentypen größer wird.

Es lassen sich abschließend folgende Empfehlungen zu den Ausgangsfragen dieser Master-Thesis aussprechen:

1. **Einsatz in deutschen Strafverfahren**

   Der Einsatz von eDiscovery-Systemen in deutschen Strafverfahren ist grundsätzlich unproblematisch da die wichtigen Informationen zum Nachweis der Authentizität (Prüfsummen, Zeitstempel, Pfadangaben, Indexinformationen) auch im Verarbeitungsergebnis unverfälscht wiedergegeben werden. Allerdings muss je nach Untersuchungsziel bei der jetzigen Qualität von eDiscovery-Programmen immer die IT-forensische Arbeit durch eine klassische IT-Forensik-Untersuchung begleitet werden, um die Differenz nicht erkannter Dateitypen zu kompensieren und auch sonstige relevante Tathandlungen zu erkennen, die mit dieser Technologie nicht herausgearbeitet werden können. Das bedeutet für den Arbeitsbereich der IT-Forensik Mehrarbeit.

2. **Differenzierter Einsatz**

   Den eDiscovery-Anbietern ist es bisher nicht gelungen, wie in der Privatwirtschaft auch in Strafverfahren alle möglichen digitalen Quellen und deren Aufbereitung unter einer eDiscovery-Anwendung zu vereinen, denn dadurch könnten die grössten Effizienzgewinne erzielt werden.

   Im Gegensatz zu vielen genutzten Standardanwendungen in der Privatwirtschaft müssen die Ermittlungsbehörden alle möglichen digitalen Spuren auf Relevanz untersuchen. Das bedeutet z.B. auch neue Spurenquellen aus dem Bereich der Internet of Things (IoT) - Gruppe, mit all seinen neuen Datenformaten untersuchen zu können. Durch die beschriebene Verarbeitungsmethodik bei eDiscovery-Systemen ist es illusorisch, zu glauben, dass es eine endliche Dateityp-Lösung geben wird. Denn mit jeder neuen Softwareanwendung mit möglicher Verfahrensrelevanz entstehen auch neue Dateitypen. Schon bei der Verarbeitung von Datentypen aus Mobilgeräten wurden große technische Hürden festgestellt, die bis heute nicht zufriedenstellend gelöst werden konnten.

   Es bleibt abzuwarten, ob es die Softwarenentwickler für die wichtigsten digitalen Spurenquellen in Strafverfahren (wozu insbesondere Mobilgerätedaten gehören) schaffen, eine Gesamtlösung zu entwickeln. Bis dahin sollten die angebotenen eDiscovery-Tools je nach ihren festgestellen Stärken eingesetzt werden. NUIX hat sich z.B. im Bereich von großen und umfangreichen Wirtschaftsdelikten bewährt,

da hier die effiziente Suche nach verfahrensrelevanten indexierten Textinhalten besonders wichtig und erfolgreich ist. PATHFINDER der Fa. Cellbrite könnte in der Zukunft eine Lösung für die Auswertung und Analyse von Mobilgerätedaten ebenfalls in Großverfahren werden, bei denen die Kommunikationsbeziehungen zwischen den Tätern eine wichtige Rolle zur Fallaufklärung spielen.

3. Aufbau einer eDiscovery-Lösung – Ressourceneinsatz vs. Vorteile

Gerade zum Anfang beim Aufbau derart komplexer Systeme und dem Sammeln von Erfahrung, um die notwendige Qualität bei den Verarbeitungsergebnissen und die Weiterverarbeitung durch KI - Analysen zu erreichen, sind zunächst Personalinvestitionen in hoch qualifizierte Spezialisten notwendig, bevor damit nachweisbare Arbeitszeiteinsparungen und Mehrwerte vor allem auf der Auswerter(Ermittler)-Seite erreicht werden können.

Allerdings kann zukünftig auf diese Technologie nicht verzichtet werden, wenn sie maßgeblich und effizient durch ihre neuen Möglichkeiten schneller und besser zur Fallaufklärung beitragen kann. Bis zum tatsächlichen Bedarf einer eDiscovery-Nutzung in einem aktuellen Großverfahren sollte nicht abgewartet werden, da der zeitliche Vorlauf für den Aufbau und Betrieb gerade für eine netzwerkgestützte Lösung erfahrungsgemäß mehrere Jahre dauert.

Der Einsatz von Electronic Discovery-Systemen in komplexen und großen Verfahren (mit vielen Tatbeteiligten) stellt nachgewiesenermaßen eine Verbesserung der Arbeitsergebnisse im Vergleich zu bestehenden Bearbeitungsmethoden dar. Immer mehr Länderpolizeien schaffen sich aus diesem Grunde derartige Systeme an oder sind gerade dabei, die bestehenden regionalen eDiscovery-Systeme zu Landeslösungen auszubauen.

Positiv-Effekte bei Personaleinsparungen ergeben sich dann vor allem auf der Auswerterseite. Personaleinsparungen bei der IT-Forensik ergeben sich erst dann, wenn sie klassische IT-Forensik-Untersuchungsarbeiten zu großen Anteilen ersetzen, was derzeit nicht der Fall ist. Aus diesem Grunde ist auch erklärbar, warum die bereits eDiscovery einsetzenden Polizeien des Bundes und der Länder neue Organisationseinheiten aufbauen oder personell ihre IT-Forensik-Einheiten mit

dieser Aufgabenstellung verstärken. Eine Einsparung des vorhandenen IT-Forensik-Fachpersonals konnte bei den nutzenden Länderpolizeien nicht festgestellt werden.

4. **Archivierung digitaler Spuren in Strafverfahren**
   Bei ausschließlichem Einsatz von eDiscovery-Anwendungen sollte nicht nur das Verarbeitungsergebnis, sondern immer auch die forensischen Sicherungen hierzu archiviert werden, um eine nachträgliche klassische IT-forensische-Untersuchung zu gewährleisten, die ggf. zur Überprüfung von Verarbeitungsergebnissen benötigt wird oder die schon angesprochene Differenz nicht verarbeiteter Dateitypen zumindest nachträglich ermöglicht.

4.2.2 Ausblick

In den USA, wo eDiscovery-Anwendungen ursprünglich entwickelt wurden, findet der Einsatz von eDiscovery mittlerweile bereits in größerer Anzahl, vor allem in komplexen Wirtschaftsstrafverfahren, statt.

Das US-Bundesjustizministerium hat beginnend mit dem Jahr 2015 angefangen, über den Einsatz von eDiscovery-Technologien in Strafverfahren zu informieren. In der Veröffentlichung *Criminal e-Discovery – A Pocket Guide for Judges* [38] wird auf verschiedene Fragestellungen im Kontext zwischen technologischen Machbarkeiten, Begrenzungen, Forderungen der Verteidigung und dem dortigen US-Strafrecht eingegangen, um so den Strafprozess führenden Bundesrichtern Handlungssicherheit zu geben.

Bisher sind keine Gerichtsverfahren bekanntgeworden, bei denen Fehler in den Verarbeitungsergebnissen in eDiscovery-Systemen durch die Verteidigung nachgewiesen wurden. Angreifbarkeiten und damit ein Beweismittelverwertungsverbot entstehen dann, wenn das eingesetzte eDiscovery-Tool nachweislich das Ausgangsmaterial nachteilig verändert und damit nicht mehr die Merkmale an forensische Arbeit erfüllt.

Nach Ansicht des Autors wird der hohe Automatisierungsgrad in der Verarbeitung und Aufbereitung von digitalen Rohdaten zunehmend klassische IT- forensische Workflows ablösen, da der Bedarf auf Untersuchung und Aufbereitung von digitalen möglichen verfahrensrelevanten Daten auch weiterhin steigen wird und dieser neue Workflow das

[38] http://fln.fd.org/files/training/2017/09/Pocket_Guide_for_Judges.pdf

Überlastungsproblem und damit die langen Wartezeiten teilweise lösen kann. Zudem bietet die Auswertung unter einer einzigen Bedienoberfläche für alle aufbereiteten digitalen Daten ein Höchstmaß an Effizienz und spart so wertvolle Arbeitszeit der Ermittler. Zwar werden eDiscovery-Lösungen nie eine 100%ge-Lösung für alle anfallenden IT-forensischen Aufgabenstellungen bieten, doch je mehr die Fortentwicklung technischer Lösungen der Anbieter eine ganzheitliche Lösung für die möglichen verfahrensrelevanten Daten bereitstellt, um so mehr Raum gewinnt seine Anwendbarkeit für unterschiedlichste IT-Forensik-Arbeiten in deutschen Strafverfahren, insbesondere unter den Vorteilen einer Vernetzung und einheitlichen alleinigen Bedien- und Auswerteoberfläche.

In der Industrie werden große technische Entwicklungssprünge, die die Effizienz bei der Erzeugung von Produkten erheblich steigern, wie z.B. die Verlagerung der Montage von Autos auf ein weiterlaufendes Produktionsband wie bei Henry Ford oder in den Schlachthöfen von Chicago Anfang des 20. Jahrhunderts, mit einer neuen Versionsnummer – Industrie 2.0 - versehen.
Obwohl eDiscovery-Anwendungen noch Makel besitzen, so sind auch dort Effizienzsteigerungen messbar und öffnen vor allem das Feld für Big Data-Analysen mit ganz neuen Erkenntnisgewinnen. Der Autor sieht darin ebenso einen Meilenstein in der technischen Fortentwicklung wie in der Industrie, welches eine neue Beschreibung hierfür rechtfertigt: IT-Forensik 2.0.

Abbildungverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

AT & T	American Telephone and Telegraph
BSI	Bundesamt für Sicherheit in der Informationstechnik
EDRM	Electronic Discovery Reference Model
EXIF	Exchangeable Image File
ISM	Informationssicherheitsmanagement
HTTP	Hypertext Transfer Protocol
IT	Informationstechnik
OCR	Optical Character Recognition
MBR	Master Boot Record
USB	Universal Serial Bus

Glossar

Attachement Ein Attachement ist eine Datei, die als Anhang mit einer E-Mail verschickt wird. Jede Art von Datei lässt sich an eine E-Mail anhängen.

Blu-ray Die Blu-ray Disc (BD meist Blu-ray abgekürzt) ist ein digitales optisches Speichermedium. Sie wurde als High-Definition-Nachfolger der DVD entwickelt und bietet ihrem Vorläufer gegenüber eine erheblich gesteigerte Datenrate und Speicherkapazität. Auf Blu-rays können daher Filme und Musik (Pure Audio) mit deutlich höherer Auflösung gespeichert werden.

Brute Force Die Brute-Force-Methode (aus dem Englischen: rohe Gewalt bzw. ausschöpfende Methode) ist eine Lösungsmethode, die auf dem Ausprobieren aller möglichen (oder zumindest vieler möglicher) Passwort-Varianten beruht.

Client-Server Das Client-Server-Modell (auch *Client-Server-Konzept*, *-Architektur*, *-System* oder *-Prinzip* genannt) beschreibt eine Möglichkeit, Aufgaben und Dienstleistungen innerhalb eines Netzwerkes zu verteilen. Die Aufgaben werden von Programmen erledigt, die in Clients und Server unterteilt werden. Der Client kann auf Wunsch einen Dienst vom Server anfordern (z. B. ein Betriebsmittel). Der Server, der sich auf demselben oder einem anderen Rechner im Netzwerk befindet, beantwortet die Anforderung (das heißt, er stellt im Beispiel die Betriebsmittel bereit); üblicherweise kann ein Server gleichzeitig für mehrere Clients arbeiten.

Compound File Ein zusammengesetzes Dateiformat zum Speichern zahlreicher Dateien und Streams in einer einzigen Datei auf einem Datenträger.

Container Datei Datei, die mehrere andere Dateien enthält, im Allgemeinen für Komprimierung oder Sicherheit.

Datenaufbereitung Unter Datenaufbereitung versteht man die Bereinigung und Transformation von Rohdaten vor der eigentlichen Verarbeitung und Analyse. Die Aufbereitung ist ein wichtiger Schritt vor der Verarbeitung und umfasst häufig das erneute Formatieren von Daten, die Berichtigung von Informationen und die Kombination von Datensätzen zur Anreicherung dieser Daten.
Die Datenaufbereitung ist häufig eine langwierige Aufgabe für Datenexperten oder Business-Anwender – trotzdem ist sie eine wichtige Voraussetzung, um einen Zusammenhang zwischen den Daten herzustellen. Nur so lassen sich wertvolle Erkenntnisse gewinnen und eine Verzerrung der Informationen aufgrund schlechter Datenqualität vermeiden.
Sie umfasst typischerweise die Standardisierung von Datenformaten, die Anreicherung von Quelldaten und/oder die Beseitigung von Ausreißern.

Dateiformat bzw. -typ Ein Dateiformat definiert die Struktur und den Aufbau von Daten innerhalb einer Datei. Über diese Funktionalität können moderne Betriebssysteme

Daten bestimmten Anwendungen zuordnen, die die Dateien interpretieren können: PDF-Datei => Adobe Acrobat Reader.

Datenexport- bzw. Ausgabeformat Zum Exportieren von Daten werden Dateien mit fester Länge erzeugt, die von anderen Programmen geöffnet oder importiert werden können. Typische Exportformate sind z.B. csv, dbf oder xml.

DVD Die DVD ist ein digitaler, optischer Datenspeicher, der im Aussehen einer CD ähnelt, aber über eine höhere Speicherkapazität verfügt. Das Akronym „DVD“ geht ursprünglich auf die Abkürzung von Digital Video Disc zurück. Heute wird die Abkürzung als Digital Versatile Disc (engl. für *digitale vielseitige Scheibe*) interpretiert.

eDiscovery Mittels eDiscovery (auch E-Discovery oder e-discovery) werden in Unternehmen für einen bestimmten Sachverhalt relevante Daten (meist E-Mails und Dokumente) identifiziert, aufbereitet und bereitgestellt bzw. an Dritte übergeben.
Der Begriff kommt ursprünglich aus dem angloamerikanischen Rechtsraum und bezeichnet dort den Teil eines Discovery, der elektronische Unterlagen, wie beispielsweise E-Akten, E-Mails oder Chat-Protokolle betrifft. Mittels eines eDiscovery-Prozesses soll dabei die Vollständigkeit der Daten sichergestellt und gleichzeitig die Gefahr, Geschäftsgeheimnisse zu verlieren, minimiert werden. Der Prozess wird in der Regel durch eine Software unterstützt bzw. durch diese teilautomatisiert.

Embedded Object Ein eingebettetes Objekt ist ein Objekt, welches separat erstellt und dann in einem anderen Objekt oder Programm platziert wird. Eingebettet Objekte sind eigenständig und können unahängig voneinander arbeiten. Eingebettet Objekte sind so konzipiert, dass sie physisch innerhalb des zusammengesetzten Objekts mit allen Informationen gespeichert werden, die für seine Verwaltung erforderlich sind.

Festplatte Ein Festplattenlaufwerk (englisch *hard disk drive*, Abkürzung HDD), oft auch als Festplatte oder Hard Disk (abgekürzt HD) bezeichnet, ist ein magnetisches Speichermedium der Computertechnik, bei welchem Daten auf die Oberfläche rotierender Scheiben (auch englisch „Platter“ genannt) geschrieben werden.

Filesharing Filesharing ist das direkte Weitergeben von Dateien zwischen Benutzern des Internets (meist) unter Verwendung eines Filesharing-Netzwerks. Dabei befinden sich die Dateien normalerweise auf den Computern der einzelnen Teilnehmer oder dedizierten Servern, von wo sie an interessierte Nutzer verteilt werden. Für den Zugriff auf Filesharing-Netzwerke sind entsprechende Computerprogramme (siehe unten) erforderlich

Instant Messanger IM sind eine Kommunikationsmethode, bei der sich zwei oder mehr Teilnehmer per Textnachrichten unterhalten. Dabei löst der Absender die

	Übermittlung aus, sodass die Nachrichten möglichst unmittelbar (instant) beim Empfänger ankommen.
IT- forensische Arbeiten	Die IT-forensische Arbeit besteht aus 3 grundlegenden Arbeitsschritten: 1. gerichtsfest Daten sichern 2. extrahieren und analysieren 3. Untersuchungsergebnisse verständlich dokumentieren und präsentieren sowie ggf. in auswertbares Format exportieren.
KPMG	Internationale Wirtschaftsprüfungsgesellschaft, dessen Firmenname sich aus den Anfangsbuchstaben der Gründer oder Vorsitzenden der fusionierten Gesellschaften zusammensetzt – Klynveld, Peat, Marwick, Goerdeler.
Prozessmodell	Ein Prozessmodell ist das Abbild eines Prozesses (von Arbeitsschritten) oder Prozesssystems und dient zur sinnvollen Ergänzung von Prozessbeschreibungen. Mit Hilfe des Prozessmodells lassen sich die wesentlichen Elemente, Eigenschaften und Schnittstellen verstehen und ggf. korrigieren und sogar komplexe Prozesse darstellen.
PWC	Internationale Wirtschaftsprüfungsgesellschaft, dessen Firmenname sich aus den Anfangsbuchstaben der Gründer oder Vorsitzenden der fusionierten Gesellschaften zusammensetzt – Price, Waterhouse, Cooper.
Rendering	Der Begriff Rendering heißt wörtlich aus dem Englischen übersetzt: Wiedergabe oder Übersetzung. Im Computerbereich ist es der Oberbegriff für Verfahren oder Software zur Wiedergabe von zweidimensionalen Pixelbildern oder von dreidimensionalen Vektor-Objekten unter Berücksichtigung von Texturen, Effekten und/oder Lichtquellen.
Social Media and Collaboration Plattform	Eine Kollaborationsplattform ist eine Software, die Mitarbeiter-Teams dabei unterstützt, bestimmte Ziele zu erreichen oder Geschäftsprobleme durch Dokumentenmanagement, Ideenaustausch und Aufgabenverwaltung zu lösen.
Workflow	Ein Workflow ist die Abwicklung arbeitsteiliger Vorgänge bzw. Geschäftsprozesse in Unternehmen und Behörden an einem Arbeitsplatz. Dabei hat der Workflow einen definierten Startpunkt, Ablauf und Endpunkt.

Literaturverzeichnis und Online-Quellen

1. Das Electronic Discovery Reference Modell
   https://edrm.net/

2. Processing in E-Discovery, a Primer von Craig Ball
   http://www.craigball.com/Ball_Processing_2019.pdf

3. E-Discovery und Computerforensik – wie unterscheiden Sie sich? von Hannah George vom 30.01.2018
   https://resources.infosecinstitute.com/topic/e-discovery-computer-forensics-different/

4. eDiscovery & Dokumentenreview – Technische Möglichkeiten beim Einsatz von eDiscovery-Werkzeugen anläßlich der IT-Fachanwaltstagung München 2019 von Dipl. Inform. Fabian Unucka, Fa. FAST DECTECT
   https://www.it-fachanwaltstage.de/assets/Muenchen/Vortraege/IT-Fachanwaltstage-2019-Vortrag-Unucka-Fast-Detect-e-discovery.pdf

5. Bereicherung von NUIX Processing und NUIX Investigate for Information Governance von Phil Glod, Senior Consultant NUIX vom 28.01.2021 veröffentlicht bei www.forensicfocus.com
   https://www.forensicfocus.com/webinars/enriching-nuix-processing-and-nuix-investigate-for-information-governance/

6. Leitfaden IT Forensik Version 1.0.1 (März 2011) des BSI
   https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Leitfaden_IT-Forensik.pdf?__blob=publicationFile&v=1

7. Welche Änderungen müssen wir in eDiscovery sehen? Teil I – VII von Harold Burt-Gerrans, Director, eDiscovery and Computer Forensics bei Epiq, Canada v. 07.10.2019 veröffentlicht bei www.forensicfocus.com
   https://www.forensicfocus.com/articles/what-changes-do-we-need-to-see-in-ediscovery-part-i

8. Computer Forensik, „4.2 Anforderungen an den Ermittlungsprozess“, S. 62 – 63 und „4.4 Das S-A-P Modell“, S. 64 -65, 3. Auflage, ISBN 978-3-89864-534-8 von Alexander Geschonneck

9. Criminal e-Discovery, A Pocket Guide for Judges, US Federal Juidicial Center von Sean Broderick, Donna Lee Elm, Andrew Goldsmith, John Haried, Kiran Raj, veröfffentlicht unter http://fln.fd.org/files/training/2017/09/Pocket_Guide_for_Judges.pdf

Thesen

- Das Prinzip von eDiscvovey-Anwendungen durch vorgefertigte Lösungen für bestimmte Dateitypen Verarbeitungsergebnisse liefern zu können, birgt das Risiko, dass Spuren, die in Dateitypen enthalten sind, die nicht unterstützt werden, auch nicht im Verarbeitungsergebnis enthalten sind.

- Electronic Discovery-Systeme können im derzeitigen Leistungsumfang bestehende klassische IT-forensische Arbeit nicht komplett ersetzen, da sonst die Gefahr besteht, dass aufgrund der Verarbeitungsmethodik Spuren im Verarbeitungsergebnis nicht mehr auftauchen. Daher muß der Einsatz von eDiscovery-Systemen auch immer durch eine klassische IT-Forensik-Arbeitsweise begleitet werden.

- Electronic Discovery-Systeme ermöglichen durch die Harmonisierung der unterschiedlichen Datenformate und Übertragung in eine oder mehrere Datenbanken den Einsatz von Big Data Analysen. Diese Mehrwerte IT -forensischer Untersuchungsarbeit sind wichtige neue Erkenntnisquellen für die Aufklärung von Straftaten.

- Bestimmte Aufgabenstellungen aus der IT Forensik sind nicht über den Einsatz von eDiscovery-Anwendungen lösbar. Hierzu gehören vor allem Fragestellungen zu Tathandlungen bzw. der Nutzung von IT - Gerät für strafbewehrte Handlungen. Dazu ist es notwendig, die Gesamtheit der Daten zu erhalten, um einzelne Spuren nachzuverfolgen, die Abhängigkeiten oder Zusammenhänge zwischen den eingesetzten Programmen zu erkennen und aus diesen Zusammenhängen erste Theorien entwickeln zu können. Bei eDiscovery-Lösungen wird nur noch mit reduzierten Datenbeständen gearbeitet, so dass durch das Fehlen dieser Informationen keine Tatzusammenhänge und auch das Vorgehen selber nicht mehr nachvollzogen werden kann.

Ehrenwörtliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Masterarbeit selbstständig geschrieben und ich die Übernahme wörtlicher Zitate aus der Literatur sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Kiel, d. 16.02.2023

Frank Meixelsperger

Anlagen

Auf der beiliegenden DVD zu dieser Master Thesis befinden sich alle im Rahmen der Tests von den eingesetzten Programmen generierten Protokolle.

Über die Protokolle werden die Programm-Versionsstände, der Zeitpunkt des Einsatzes, und deren Verarbeitung und Ergebnisse dokumentiert, so daß die in der Master Thesis erklärten Untersuchungsergebnisse in ihrer Gesamtheit auch nachträglich nachvollziehbar sind.

Inhalt:

Master-Thesis Frank Meixelsperger im PDF-Format

AXIOM 6.8 – Protokollsammlung zu den beiden Images für Test 2

NUIX 8.8 – Protokollsammlung zu den beiden Images für Test 2