Hochschule Wismar
Fakultät für Ingenieurwissenschaften
Bereich Elektrotechnik und Informatik

Hochschule Wismar

Fakultät für Ingenieurwissenschaften

Durchführung forensischer Datenanalysen unter Verwendung von interaktiven Grafiken

Masterarbeit zur Erlangung des Grades eines

Master of Engineering

der Hochschule Wismar

eingereicht von: Patrick Thoma
Studiengang: IT-Sicherheit und Forensik

Erstgutachterin: Prof. Dr.-Ing. Antje Raab-Düsterhöft
Zweitgutachter: Prof. Dr. Matthias Kreuseler

Aufgabenstellung

Das Ziel dieser Arbeit ist es, herauszufinden, unter Verwendung welcher Methoden und welcher nutzerfreundlichen Mensch-Technik-Schnittstellen es möglich ist, forensische Datenanalysen mithilfe interaktiver Grafiken durchzuführen. Um diese Forschungsfrage zu beantworten, werden bestehende Methoden und Werkzeuge zur forensischen Datenanalyse erfasst und bewertet. Zur fallbezogenen Bestimmung verschiedener Darstellungsmöglichkeiten wird ein mehrstufiges Auswahlverfahren definiert. Dieses ermöglicht die Visualisierung von gesicherten Datenbeständen, auf Basis von zuvor erfassten Anforderungen. Durch die Implementierung eines Falles, der auf Testdaten basiert, werden die Funktion des Auswahlverfahrens und die daraus resultierenden Ergebnisse überprüft. Die Klassifikation der infrage kommenden Datentypen wird hierzu auf Text sowie ein- und multidimensionale Daten mit einem Zeitbezug beschränkt. Diese Daten weisen ein strukturiertes, semi- oder unstrukturiertes Format auf. Nach der Beschreibung von möglichen Einschränkungen und Herausforderungen bei der Daten- und Informationsvisualisierung, im Hinblick auf den Einsatz während einer forensischen Untersuchung, wird die Arbeit mit einem Ausblick abgeschlossen.

Task definition

The aim of the work is to determine which methods and user-friendly human-machine interfaces can be used to perform forensic data analysis using interactive graphics. To answer this research question, existing methods and tools for forensic data analysis are collected and evaluated. A multi-stage selection procedure is defined for the case-related determination of different presentation options. This enables the visualization of stored data based on previously recorded requirements of forensic data analysis for information visualization. By implementing a case that is based on test data, the function of the selection procedure and the outcoming results are reviewed. The necessary classification of the data types in question is limited to text as well as one- and multidimensional data with a time reference. This data is in a structured, semi-structured or unstructured format. After the description of possible limitations and challenges in data and information visualization related to the use during a forensic investigation, the work is concluded with an summary and outlook.

Beim Erstellen dieser Arbeit wird aus Gründen der besseren Lesbarkeit auf die zusätzliche Formulierung der weiblichen Form verzichtet. Die ausschließlich männliche Form soll als geschlechtsneutral verstanden werden.

Abstract

Es existieren verschiedene Methoden, um Daten und Informationen forensisch zu analysieren. Diese Vielfältigkeit macht es möglich, digitale Spuren, die häufig mit einer illegalen oder schadhaften Handlung einhergehen, in einer forensischen Untersuchung zu entdecken. Um komplexe und umfangreiche Vorfälle effektiv zu behandeln, ist es notwendig, dass relevante Ereignisse sicher identifiziert und im Kontext mit anderen Informationen dargestellt werden können.

Das Ziel dieser Arbeit ist es zu bestimmen, wie die Darstellung von Daten und Informationen in Form von interaktiven Grafiken in der forensischen Datenanalyse angewendet werden kann. Hierzu wird auf Grundlage einer Fall- und Ereignisklassifikation eine Menge an verfügbaren Methoden zur Datenvisualisierung bestimmt. Diese werden in ein Auswahlverfahren überführt. Das Auswahlverfahren beschreibt geeignete Visualisierungen zur Darstellung von zehn Fall- und Ereignisklassen. Durch die, auf dem Auswahlverfahren basierende, Implementierung eines Testfalls ist es dann möglich, den Einsatz von interaktiven Grafiken zur forensischen Datenanalyse zu bewerten.

Das Verfahren und die im Verlauf aufgezeigten Möglichkeiten können dabei unabhängig von der Fall- oder Ereignisklasse weiterentwickelt werden, sodass es möglich ist, die verfügbaren Visualisierungsmethoden bestmöglich einzusetzen und so die Analysetätigkeit nachhaltig zu unterstützen.

Abstract

There are different methods to analyze data and information using forensic methods. This diversity makes it possible to detect digital traces, often associated with an illegal or harmful act, in a forensic investigation. In order to effectively handle complex and large incidents, it is necessary that relevant events can be reliably identified and contextualized with additional information.

The aim of this thesis is to determine how the representation of data and information in the form of interactive graphics can be applied in forensic data analysis. For this purpose, a set of available methods for data visualization is determined based on a case and event classification. These are then used in a selection procedure. This selection procedure describes suitable visualizations for representing ten case and event classes. By implementing a test case based on the selection procedure, it is then possible to evaluate the use of interactive graphics for forensic data analysis.

The procedure and the possibilities shown in the following can be enhanced independently from the case or event classes, so that it is possible to use the available visualization methods in the best possible way and thus support the analysis activity in a sustainable way.

Inhaltsverzeichnis

1 Einleitung ... 1
1.1 Motivation ... 1
1.2 Ziele der Arbeit ... 3
1.3 Nutzen und Zielgruppe: Partizipation als Schlüssel ... 4
1.4 Abgrenzung: Visualisierung in der forensischen Datenanalyse ... 5
1.5 Gliederung der Arbeit ... 5
1.6 Verwandte Arbeiten ... 6
2 Problembeschreibung und Forschungsfrage ... 7
3 Die forensische Datenanalyse ... 8
3.1 Entstehung der digitalen Forensik ... 8
3.2 Grundlagen der forensischen Datenanalyse ... 10
3.2.1 Allgemeine Definition und Begriffsabgrenzung ... 12
3.2.2 Vorgehen und Ziel ... 13
3.2.3 Erfolgsbedingungen ... 16
3.3 Einsatzgebiete der forensischen Datenanalyse ... 19
3.3.1 Zivil- und strafrechtliche Untersuchung ... 19
3.3.2 Technische Fehleranalyse ... 21
3.4 Herausforderungen vor und während der Datenanalyse ... 23
4 Die Daten- und Informationsvisualisierung ... 25
4.1 Hintergrund und Entstehungsgeschichte ... 26
4.2 Grundlagen der Daten- und Informationsvisualisierung ... 28
4.2.1 Allgemeine Definition ... 29
4.2.2 Nutzen der Datenvisualisierung ... 29
4.2.3 Erstellung effektiver Visualisierungen ... 32
4.3 Einsatzgebiete der Daten- und Informationsvisualisierung ... 34
4.3.1 Log Management ... 35
4.3.2 eDiscovery ... 37
5 Visualisierungsmethoden und Werkzeuge zur forensischen Datenanalyse ... 42
5.1 Informationsvisualiualisierung als Hilfsmittel während der forensischen Datenanalyse ... 43
5.2 Verwendungskontext der Methoden und Werkzeuge ... 44
5.2.1 Ereignisprotokolle ... 45
5.2.2 Verbindungs- und Geodaten ... 47
5.2.3 Dateisysteminformationen ... 48
5.2.4 Textdaten ... 50
5.3 Funktion und Nutzen eingesetzter Methoden und Werkzeuge ... 51

6 Anforderungen der forensischen Datenanalyse an die Daten- und Informationsvisualisierung 53
6.1 Allgemeine Anforderungen an die IT-Forensik 53
6.2 Organisatorische Anforderungen 55
6.3 Technische Anforderungen 57
6.4 Ergebnisdokumentation 58
7 Auswahlverfahren zur Bestimmung der Visualisierungsmethoden 60
7.1 Motivation 60
7.2 Modellbildung 61
7.2.1 Modellgrundlagen 62
7.2.2 Fall- und Ereignisklassifikation 63
7.2.3 Bestimmung der Visualisierungsmethode 74
7.3 Auswahlverfahren 83
8 Anwendung des Auswahlverfahrens zur Visualisierungserstellung 96
8.1 Fallbestimmung und Datenstruktur 96
8.2 Implementierung des Fallbeispiels 97
8.3 Visualisierung der Vorfallsdaten 101
9 Einschränkungen und Herausforderungen bei der Daten- und Informationsvisualisierung 111
10 Fazit 113
10.1 Zusammenfassung 113
10.2 Bewertung 114
10.3 Ausblick 115
Abkürzungsverzeichnis 117
Abbildungsverzeichnis 118
Tabellenverzeichnis 122
Quellen- und Literaturverzeichnis 123
Thesen 133
Selbstständigkeitserklärung 134

1 Einleitung

In dieser Arbeit wird der Einsatz von interaktiven Grafiken innerhalb der Datenanalyse, die ein fester Bestandteil einer jeden forensischen Untersuchung ist, beschrieben. Zu diesem Zweck beinhalten die folgenden Unterkapitel die Motivation, die Ziele sowie den Nutzen und die Zielgruppe dieser Form der Daten- und Informationsanalyse.

1.1 Motivation

Der Beginn des Digitalzeitalters im Jahr 2002 hat das gesellschaftliche Zusammenleben in weiten Teilen verändert. Allgemein sicht- und spürbar werden diese Veränderungen im Gesellschafts-, Wirtschafts- und Privatleben. Da dies ausnahmslos für alle Umbrüche innerhalb der Menschheitsgeschichte gilt, ist zu bestimmen, in welchen Teilen diese Entwicklungen Handlungsbedarf hervorrufen und wo sie durch bestehende Normen und Konventionen das Gesellschafts-, Wirtschafts- und Privatleben einzelner Personen oder Gruppen, sowohl positiv als auch negativ, kaum beeinflussen (Stengel et al. 2017, S. 17-19). Die Errungenschaften in den Bereichen Automatisierung, Vernetzung und Mobilität führen indes auch dazu, dass materielle Verbrechen differenzierter und damit im Vergleich zur bisher bekannten Massenkriminalität auch subtiler begangen werden können (BKA 2003, S. 3-35).

Getragen durch diese Entwicklungsgeschichte und die damit verbundenen Herausforderungen bei der Aufklärung dieser Tatbestände wurden bestehende Disziplinen der forensischen Wissenschaft in Inhalt und Umfang angepasst. Bewährte Verfahren und Methoden wurden durch die Technisierung vereinfacht, verändert und auch ersetzt (McDermid 2016, S. 9-33). Die Zielstellung aller Disziplinen der forensischen Wissenschaft, unabhängig von den Einflüssen und Veränderungen des digitalen Zeitalters, bleibt jedoch unverändert die Bestimmung von objektiven Befunden und Spuren. Gestützt wird dieses Ziel durch das Locard'sche Prinzip, das bis heute für die Kriminalistik und damit auch für die Forensik und deren Arbeitsgebiete von zentraler Bedeutung ist. Das von Dr. Edmond Locard formulierte Austauschprinzip – jeder, d. h. der Täter und das Opfer, und alles, also auch das Tatmittel, lassen am Tatort etwas zurück

oder nehmen etwas von dort mit – bildet die Grundlage für ein Vorgehen zur Bestimmung von Hauptbelastungsbeweisen (Locard 1930). Dieses Austauschprinzip, das die Verbindung zwischen dem Tatort, dem Opfer, den Spuren und dem Verdächtigen beschreibt, ist in Abbildung 1 dargestellt.

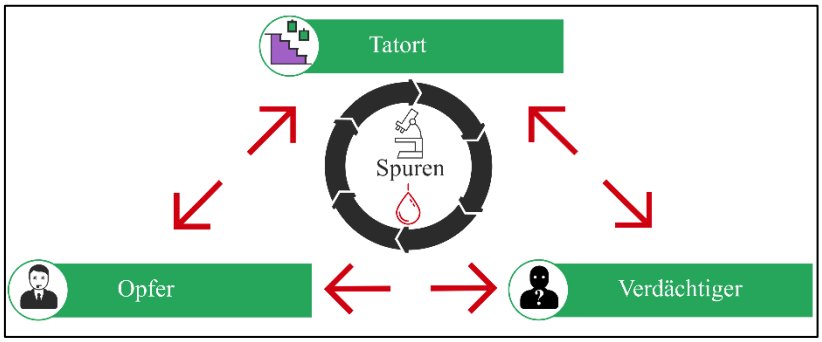

Abbildung 1: Austauschprinzip nach Dr. Edmond Locard (Eigene Darstellung, in Anlehnung an: (Locard 1930))

Diese Verbindungen können unabhängig von der straf- oder zivilrechtlichen Bedeutung des Ereignisses zur Beschreibung und Aufklärung von Vorfällen genutzt werden. Neben bestehenden Methoden der Geistes- und Naturwissenschaften werden in der forensischen Wissenschaft auch Vorgehensweisen und Techniken zur Analyse von digitalen Aufzeichnungen verwendet. Diese digitalen Spuren können Daten sein, die in einem Computersystem gespeichert oder durch dieses übertragen worden sind (Casey 2011, S. 187-224). Der Einsatz dieser Methoden und Praktiken soll die Rekonstruktion von Vorfällen auf Basis gesicherter Spuren ermöglichen. Anwendungsfälle sind die Auswertung von Bild- und Tonaufnahmen sowie die visuelle Darstellung von Täter, Opfer, Tatwerkzeug und Tatort. Das immanente Ziel aller verfügbaren Methoden, Techniken und Vorgehensweisen ist es, die ermittelten Erkenntnisse nachvollziehbar und verständlich aufzubereiten, sodass sowohl Experten als auch fachfremde Gutachter oder Beteiligte einen Vorfall sowie dessen Bestandteile, Abhängigkeiten und Auswirkungen erfassen und bewerten können. Innerhalb der forensischen Wissenschaft gibt es Arbeits- und Untersuchungsgebiete, in denen physische Beweismittel existieren. Beispiele hierfür sind unter anderem die Rechtsmedizin, die Ballistik oder die Entomologie. Zudem gibt es Arbeitsgebiete wie die forensische Psychiatrie oder die IT-Forensik, in denen eine Bewertung von Vorfällen auf Basis von Beobachtungen, Verhaltensweisen oder digitalen Aufzeichnungen vorgenommen wird.

Die sich aus dem zur Beurteilung gewählten Vorgehen ergebenden Herausforderungen hinsichtlich der Nachvollziehbarkeit und Objektivität sollten berücksichtigt werden (Katz und Halámek 2016, S. 1-7). Dies gilt für alle von Spezialisten aus dem jeweiligen Fachbereich auszuführenden Tätigkeiten und ist ggf. nicht nur auf die Analyse und Bewertung von Ereignissen beschränkt, sondern umfasst auch vor- und nachgelagerte Tätigkeiten. Die Analysetätigkeit kann in Abhängigkeit von der fachlichen Disziplin ein weitestgehend manueller und damit zeitintensiver Prozess sein. Durch ein Anwachsen der Daten- und Informationsbestände werden vor allem jene Fachbereiche beeinflusst, in denen die Heterogenität der Tatbestände dazu führt, dass die inhaltliche und thematische Erfahrung von Experten mitunter nur unzureichend eingesetzt werden kann oder durch ressourcenintensive vor- und nachbereitende Maßnahmen ergänzt werden muss.

1.2 Ziele der Arbeit

In dieser Arbeit wird eine Möglichkeit aufgezeigt, wie zentrale Herausforderungen, die vor allem im Rahmen der forensischen Datenanalyse anfallen, mitigiert werden können. Losgelöst von vorgelagerten Teilprozessen innerhalb der Vorfallsbearbeitung, wie der operativen Vorbereitung oder der mitunter zeitkritischen Datensicherung, wird eine methodische Untersuchung von kriminellen Handlungen, die in einem informationstechnischen System oder durch ein solches System begangen wurden, ermöglicht. Die diesbezüglich angewandte Systematik ist streng methodisch und ermöglicht es, das gewählte Vorgehen jederzeit nachweisbar begründen zu können. Erkenntnisse, die aus der Auswertung hervorgehen, erfüllen stets zentrale Anforderungen an die Nachvollziehbarkeit und Zielorientierung der forensischen Untersuchung. Möglich wird dies, da alle durchgeführten Verarbeitungs- und Analyseschritte umfassend dokumentiert werden und damit reproduzierbar sind. Die Antwort auf die Frage ‚Welches Ereignis ist wann, wie und wo passiert?‘ bildet dabei die Grundlage für weiterführende Fragestellungen nach dem Urheber einer Handlung und mögliche reaktive, und nachgelagert auch präventive, Maßnahmen. Zu berücksichtigen sind hierbei Herausforderungen, die durch eine zunehmende Interoperabilität von System- und Anwendungskomponenten entstehen. Diese führen dazu, dass im Rahmen einer forensischen Untersuchung immer mehr Datenquellen erforderlich sind. Die sich daraus ergebende Komplexität hinsichtlich der Qualität und des Umfangs der Datenbestände

führt dazu, dass der Adressatenkreis, der mit der weiterführenden Analyse fachlich bedacht werden kann, ein umfassendes technisches Expertenwissen benötigt. Dies gilt unter anderem für große oder unstrukturierte Datenmengen.

Die Durchführung forensischer Analysen unter Verwendung von interaktiven Grafiken zur Visualisierung erfasster Datenbestände bietet dabei weitreichende Möglichkeiten. Der Analyst oder Ermittler muss dazu frei von systemspezifischen Eigenschaften der Quellsysteme agieren können. Ziel ist es, dass der Bedarf an spezifischem Wissen über die Datenquellen und deren Anforderungen weitestgehend entfällt, sodass der Personenkreis, der mit der Fallbearbeitung betraut werden kann, erweitert wird. Der Fokus bei der Fallbearbeitung verschiebt sich infolgedessen vollständig auf die forensische Untersuchung. Werden bei der Erstellung eines entsprechenden Vorgehensmodells die fachlichen Anforderungen an die operativen Fallbearbeiter berücksichtigt, ist es möglich, aufkommende Fragestellungen zentral zu beantworten und diese anlassbezogen weiterzuentwickeln. Dies ermöglicht es, einen hohen Standardisierungsgrad bei der Durchführung forensischer Datenanalysen unter Verwendung von interaktiven Grafiken zu erreichen. Eine vollständige Übersicht, aus der ein Fallbearbeiter die, für die vorhandene Datenstruktur, geeignetste Visualisierung ableiten kann, reduziert zusätzlich zeitliche und personelle Aufwände bei der fallbasierten Bestimmung der Visualisierungsmöglichkeiten. Perspektivisch werden Datenbestände so zielgruppengerecht, d. h. allgemein verständlich, aufbereitet, sodass die inhaltliche Analyse der Daten in den Vordergrund rückt und technische Hürden bei der Erstellung und dem Betrieb von Hilfsmitteln, den Visualisierungstechniken, reduziert werden.

1.3 Nutzen und Zielgruppe: Partizipation als Schlüssel

Der Beitrag dieser Arbeit liegt in der Klassifikation von Möglichkeiten der Daten- und Informationsvisualisierung im Kontext der forensischen Datenanalyse. Hierzu wird das Verhalten der beschriebenen Visualisierungstechniken im Grenzbereich, bei unvollständigen Datengrundlagen und volumenbedingten Darstellungsgrenzen, bestimmt. Der gesamte Entwurf folgt einer deduktiven Argumentation, d. h. das beschriebene Vorgehen basiert auf bestehenden Theorien der forensischen Datenanalyse und der Daten- und Informationsvisualisierung, in deren Verlauf auf spezielle Fallarten,

deren Aufbereitung und Lösungsansätze eingegangen wird. Die daraus gewonnenen Erkenntnisse, die in den Bereich der Forensigraphie fallen, werden nach Verfahren, den entsprechenden Abhängigkeiten und Anwendungsgebieten klassifiziert. Unter diesen Voraussetzungen ist es möglich, die beschriebenen Verfahren, ihre Nutzbarkeit und Optimierungspotenziale einzuordnen. Das ermöglicht die aktive Partizipation Dritter, die dieses Vorgehensmodell im Rahmen der forensischen Analyse von Vorfällen nutzen, anpassen und weiterentwickeln möchten. Der Nutzen der Arbeit liegt somit in der Beschreibung eines Werkzeugs zur Unterstützung und Durchführung von forensischen Datenanalysen durch Techniken und Methoden zur Informationsvisualisierung, das sowohl deskriptiv als auch explorativ eingesetzt werden kann.

1.4 Abgrenzung: Visualisierung in der forensischen Datenanalyse

Neu- und Weiterentwicklungen im Bereich der Informationstechnologie, der damit verbundene Fortschritt und die damit verbundenen Veränderungen beeinflussen die digitale Forensik. Die erhöhte Systemleistung, eine stärkere gesellschaftliche Durchdringung und wachsende Datenmengen von möglichen Quellsystemen haben dazu geführt, dass der forensische Prozess und die im zeitlichen Ablauf zu erbringenden Teilaufgaben neuen, teils komplexen Herausforderungen unterliegen. Eine konsolidierte Bewertung dieser Abhängigkeiten über alle Teilaufgaben des forensischen Prozesses hinweg entspricht nicht dem Ziel dieser Arbeit. Der Fokus der Arbeit liegt vielmehr auf dem Potenzial von Informationsvisualisierungstechniken zur Unterstützung und Durchführung der Datenanalyse und Untersuchung. Hierzu werden die Genauigkeit der zur Verfügung gestellten Visualisierungen und damit verknüpft die Möglichkeiten beschrieben, mit denen Analysten, Gutachter und Ermittler Schlussfolgerungen und relevante Erkenntnisse aus den dargestellten Daten ableiten können. Eine exakte Darstellung vorfallsrelevanter Ereignisse bildet hierbei die Zielstellung.

1.5 Gliederung der Arbeit

Diese Arbeit besteht aus zehn Kapiteln. In den Kapiteln 1 bis 4 werden sowohl die Grundlagen der forensischen Datenanalyse als auch die der Daten- und Informationsvisualisierung beschrieben. Die Kapitel 5 und 6 beinhalten bereits

bestehende Analysemethoden und -werkzeuge im Rahmen einer forensischen Untersuchung und die Anforderungen, die vonseiten der forensischen Datenanalyse an die Daten- und Informationsvisualisierung gestellt werden. Die Erstellung eines Auswahlverfahrens auf Grundlage einer Fall- und Ereignisklassifikation sowie die praktische Erprobung dieses Modells anhand eines Testvorfalls werden in den Kapiteln 7 und 8 beschrieben. Im Anschluss daran werden in Kapitel 9 Einschränkungen und Herausforderungen bei der Daten- und Informationsvisualisierung genannt und erläutert. Ein Fazit, bestehend aus einer Zusammenfassung, einer Bewertung und einem Ausblick, schließt die Arbeit in Kapitel 10 ab.

1.6 Verwandte Arbeiten

Die Überprüfung der verfügbaren Literatur hat gezeigt, dass die Visualisierung von Daten und Informationen innerhalb des forensischen Prozesses von zentraler Bedeutung ist. Die Autoren Labudde und Spranger beschreiben in ‚Forensik in der digitalen Welt' die Einsatzmöglichkeiten der Datenvisualisierung, unter anderem den Einsatz zur Textforensik und Tatortrekonstruktion (Labudde und Spranger 2017). In den Ausarbeitungen ‚Computer-Forensik' (Geschonneck et al. 2014) und ‚PC-Forensik' (Willer 2012) werden bestehende Möglichkeiten zur Visualisierung von Daten und Informationen, die in Dateisystemen gespeichert werden, erläutert. Neben diesen Möglichkeiten beschreiben die Autoren jedoch auch Einschränkungen in Bezug auf die Integration anderer Datenquellen, mit der vorfallsrelevante Informationen über einzelne Datenquellen hinweg verknüpft werden sollen. Die Bedien- und Anpassbarkeit der Anwendungen und Methoden zur Datenvisualisierung werden hierbei als Erfolgsfaktoren angeführt. So bieten vor allem Lösungen, die es dem Anwender ermöglichen, den Darstellungsinhalt bei Bedarf anzupassen, einen Mehrwert innerhalb der Analysephase. Eine Anwendung, die es ermöglicht, Webprotokolle für forensische Analysen grafisch aufzubereiten, wird in ‚Web History Visualisation for Forensic Investigations' (Lowman 2010) beschrieben. Eine Anpassung der verfügbaren Visualisierungsmöglichkeiten durch den Anwender ist bei dieser Lösung nicht vorgesehen. Trotz dieser Tatsache bietet die Anwendung für die Untersuchung des Browserverlaufs umfangreiche Analysemöglichkeiten und ermöglicht einen Ausblick auf die Nutzung und den Mehrwert von interaktiven Grafiken zur Datenanalyse.

2 Problembeschreibung und Forschungsfrage

Um Daten und Informationen zu einem Vorfall effektiv forensisch untersuchen zu können, werden forensische Methoden und Anwendungen eingesetzt. Diese unterstützen insbesondere die Untersuchungs- und Analysephase. Aktuell besteht keine vollständige Übersicht, die es einem Fallbearbeiter ermöglicht, die für die vorhandene Datenstruktur geeignetste Visualisierung zur Analyse eines Ereignisses abzuleiten. Das führt mitunter dazu, dass Datenbestände nicht oder nur unzureichend visualisiert werden können. Dieser Umstand beschränkt die Auswahl der Fallbearbeiter auf jene mit umfassendem technischen Expertenwissen. Die Expertise von Sachverständigen, die in der Domäne tätig sind, der ein Vorfall zuzuordnen ist, ist bei der Aufklärung häufig von zentraler Bedeutung. Die Bedienbarkeit der zur forensischen Analyse eingesetzten Methoden und Anwendungen bildet dabei die Schnittstelle zwischen der fachlichen und der technischen Expertise. Forensische Werkzeuge, deren Nutzung aufgrund ihres textuellen Aufbaus eine Einweisungs- und Ausbildungsphase vorausgeht, stellen eine Nutzungsbarriere dar, die fachfremden Analysten und Ermittlern den Einsatz dieser Werkzeuge wesentlich erschwert.

Die Forschungsfrage, die im Rahmen dieser Arbeit beantwortet werden soll, ist daher die folgende: Unter Verwendung welcher Methoden und welcher nutzerfreundlichen Mensch-Technik-Schnittstellen ist es möglich, forensische Datenanalysen mithilfe von interaktiven Grafiken durchzuführen?

3 Die forensische Datenanalyse

Eine streng methodisch vorgenommene Untersuchung von Daten ermöglicht die Aufklärung und den Nachweis von Straftaten, die in einem informationstechnischen System oder durch ein solches System begangen wurden. Im Rahmen dieses Kapitels werden daher zentrale Aspekte der forensischen Datenanalyse erläutert. Neben den fachlichen und inhaltlichen Grundlagen werden auch die Einsatzgebiete der forensischen Datenanalyse betrachtet. Hierzu werden notwendige vor- und nachbereitende Maßnahmen genannt, sodass es möglich ist, deren Einfluss auf die Phasen der Untersuchung und Datenanalyse im Rahmen des forensischen Prozesses, nach dem Bundesamt für Sicherheit in der Informationstechnik (BSI), zu beschreiben. Von wesentlicher Bedeutung sind hierbei Maßnahmen, die es ermöglichen, im Rahmen der komplexen Bewertung von Straftaten in IT-Systemen existierende Zusammenhänge ausreichend und präzise sichtbar zu machen. Dies gilt insbesondere vor dem Hintergrund, Analysten und Ermittler in die Lage versetzen zu wollen, komplexe oder große Datenmengen intuitiv analysieren zu können.

3.1 Entstehung der digitalen Forensik

Die ersten Computer (Rechenmaschinen) wurden ausschließlich durch speziell geschulte Personen bedient. Diese zählten zu einem ausgewählten Personenkreis und waren Angestellte oder Mitglieder von akademischen Einrichtungen, Regierungsorganisationen und Unternehmen. Die dort eingesetzten Rechenmaschinen waren physisch isoliert und besaßen definierte Schnittstellen zu Ein- und Ausgabegeräten. Der Datenaustausch zwischen zwei Rechenmaschinen über ein Netzwerk war zu diesem Zeitpunkt nicht möglich. Damit war es unwahrscheinlich, aber dennoch nicht unmöglich, dass diese Maschinen und ihre Rechenleistung missbraucht wurden, um dem Nutzer des Geräts einen unberechtigten Vorteil zu verschaffen oder einen unbeteiligten Dritten zu benachteiligen. Trotz der Fokussierung auf den positiven Nutzen dieser Geräte bestand die Notwendigkeit, ausgeführte Aktionen und deren Auswirkungen nachvollziehen zu können (Pollitt 2010, S. 3-15).

1976 wurden daher erstmals Möglichkeiten aufgezeigt, wie Nachweise in einem Großrechner gesammelt werden können, um diese als Beweis vor Gericht anzubringen (Parker 1976). Diese Nachweise beschränkten sich auf Daten, die im Rahmen von Revisionstätigkeiten entnommen und dann in Form einer Feststellung dokumentiert wurden. Die Informationen beschrieben damit stets eine Abweichung vom definierten Soll-Zustand und waren nicht ausschließlich auf dolose Handlungen zurückzuführen, sondern auch auf Anwendungsfehler und ungeplante Systemunterbrechungen. Der Kreis der Verdächtigen, die für die Abweichung verantwortlich sein konnten, war zu diesem Zeitpunkt also übersichtlich. Das lag vor allem daran, dass die Rechnersysteme über keine echten Schnittstellen zu anderen Systemen verfügten und die Personen, die fähig waren, die Systeme zu bedienen, eine Minderheit darstellten. Die Verfolgung von Straftaten, die in diesen Systemen oder durch diese Systeme begangen wurden, war zu diesem Zeitpunkt damit auch für Strafverfolgungsbehörden problematisch, da die Grundlagen der Bedienung solcher Rechenmaschinen komplex waren, was die Anzahl der dazu fähigen Ermittler einschränkte. Die Sicherheit und im Nachgang auch die Erkennung von Manipulationen an diesen Maschinen war für viele Anlagenbetreiber in diesem frühen Entwicklungsstadium nur von nachgelagertem Interesse. Der Fokus der Betreiber lag in der Effizienzsteigerung von Tätigkeiten und Prozessen, die zuvor manuell durchgeführt wurden. Eine der ersten Ermittlungen in einem Rechnernetz, die das Ziel hatte, den Urheber einer Handlung zu identifizieren, wurde 1989 in einem Sachbuch dokumentiert (Stoll 1989). Hintergrund dieser Untersuchung war die Ursachenforschung hinsichtlich einer Differenz von 75 Cent, die in zwei Buchhaltungsanwendungen festgestellt wurde. Im Verlauf der Untersuchung konnten Aktivitäten einer Hackergruppe identifiziert werden, die Zugang zu mehreren Computersystemen, darunter auch denen von Militäranlagen, erlangt hatten.

Mit der Weiterentwicklung des Computers und der damit einhergehenden allgemeinen Verbreitung stieg auch die Wahrscheinlichkeit von Ereignissen, die aus der Fehlbedienung durch den Nutzer bzw. dem gezielten Missbrauch von Systemfunktionen hervorgehen. Die International Organisation on Computer Evidence (IOCE) war infolgedessen eine der ersten Organisationen, die sich ausschließlich mit Themen der digitalen Forensik befasste (Forensic Science Communications 2000). Ihre Gründungsmitglieder waren Mitarbeiter verschiedener US-Behörden, darunter das Verteidigungsministerium, der Secret Service und das Federal Bureau of Investigation

(FBI). Im Jahr 1993 wurde dann durch das FBI die erste internationale Konferenz zu Computerbeweisen (engl. computer evidences) veranstaltet. An dieser Veranstaltung nahmen Vertreter aus 26 Ländern teil.

Bei Vorfällen, die zum Zeitpunkt der Konferenz diskutiert und untersucht wurden, lag der inhaltliche Fokus auf der Datenwiederherstellung. Aufgrund der Kapazitätsbegrenzungen der damaligen Speichertechnologie war es nicht selten der Fall, dass Daten, die zur Bearbeitung des Vorfalls relevant waren, überschrieben wurden. Das Internet und damit eine weitreichende Vernetzung der Computersysteme waren zu diesem Zeitpunkt nicht geläufig, sodass Straftaten in vernetzten Systemen nicht alltagsbestimmend und damit nicht Teil der Konferenz waren (Pollitt 2010, S. 3-15). Durch die Zunahme von digitalen forensischen Untersuchungen haben Forscher und Ermittlungsbehörden kontinuierlich Maßnahmen, Prozesse und Vorgaben etabliert, die eine Formalisierung des forensischen Untersuchungsprozesses fördern. Infolgedessen wurden von verschiedenen Institutionen unterschiedliche Vorgehensmodelle veröffentlicht. Diese unterscheiden sich in Bezug auf die Anwendungsgebiete und die fachliche Tiefe der Beschreibung. Im Kern sehen jedoch alle verfügbaren Untersuchungsprozesse die Phasen der Sammlung, Prüfung, Analyse und Dokumentation vor. Die chronologische Durchführung dieser Teilprozesse ermöglicht eine strukturierte forensische Untersuchung und wird daher bis heute, wenn auch in adaptierter Form, genutzt. Anpassungen der Vorgehensweise sind häufig auf Methoden innerhalb der einzelnen Phasen, die im Umfang der auszuführenden Tätigkeiten granular erweitert werden, sowie vor- und nachbereitende Maßnahmen zur Prozessunterstützung beschränkt.

3.2 Grundlagen der forensischen Datenanalyse

Das vom Bundesamt für Sicherheit in der Informationstechnik (BSI) definierte Vorgehensmodell zur Durchführung einer forensischen Untersuchung besteht aus sieben Teilprozessen. Die Ausführung der jeweiligen Teilprozesse erfolgt dabei grundsätzlich sequenziell. In Abbildung 2 ist der Prozessverlauf der forensichen Untersuchung dargestellt.

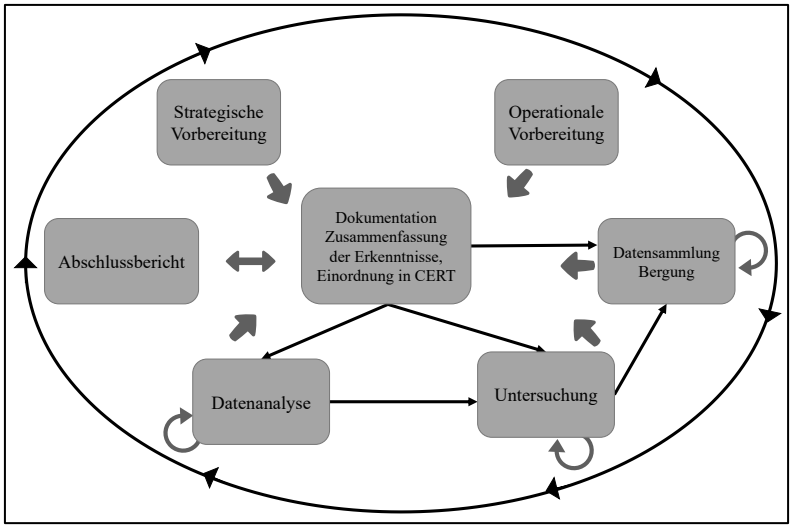

Abbildung 2: Der forensische Prozess nach BSI (Eigene Darstellung, in Anlehnung an: (BSI 2011, S. 84))

Auffällig ist, dass alle Phasen über den Teilprozess ‚Dokumentation, Zusammenfassung der Erkenntnisse, Einordnung in CERT' verbunden sind und einzelne Phasen iterativ wiederholt werden können (BSI 2011, S. 84-91). Die Abkürzung ‚CERT' steht in diesem Kontext für ‚Computer-Emergency-Response-Team'. Die Einordnung der Erkenntnisse erfolgt über die in Abbildung 3 dargestellte CERT-Taxonomie.

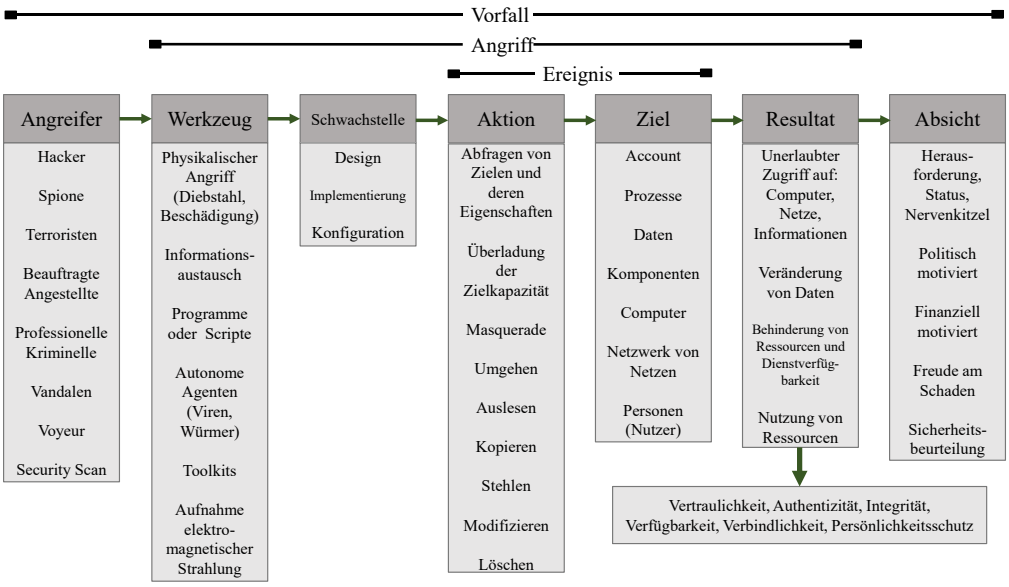

Abbildung 3: CERT-Taxonomie (Eigene Darstellung, in Anlehnung an: (BSI 2011, S. 29))

Eine Iteration über den gesamten forensischen Prozess nach dem BSI ist ebenfalls möglich. Das wiederholte Durchlaufen des forensischen Prozesses kann dann notwendig

werden, wenn die aufgrund des methodischen Vorgehens zur Aufklärung von Straftaten ausgeführten Aktionen neue Erkenntnisse und Ansatzpunkte hervorbringen. Um den Ansprüchen an ein streng methodisches Vorgehen gerecht zu werden, ist es notwendig, die ausgeführten Aktionen und entsprechende Beweggründe jederzeit nachweisen zu können (BSI 2011, S. 89-91). Aus dieser Notwendigkeit ergeben sich ebenfalls mögliche Wiederholungen im Gesamtprozess, um eben diese Informationen und deren Hintergrund zu dokumentieren. Der forensische Prozess nach dem BSI stellt eine Methodik dar, die es ermöglicht, Vorfälle in informationstechnischen Systemen durch die gezielte Analyse von vorhandenen Datenbeständen zu den beteiligten Systemkomponenten aufzuklären. Im Kern des Gesamtprozesses stehen die Untersuchung und die forensische Analyse der Datenbestände, die im Rahmen der Datensammlung geborgen werden. Daher werden die Definition, der Nutzen, mögliche Erfolgsbedingungen sowie das Vorgehen im Zuge der Analysetätigkeit in den folgenden Unterkapiteln näher erläutert.

3.2.1 Allgemeine Definition und Begriffsabgrenzung

Die forensische Datenanalyse ist ein zentraler Bestandteil der IT-Forensik. Die methodisch vorgenommene Datenanalyse von gesicherten Datenbeständen und in Computernetzwerken ermöglicht es, Thesen, die in vorgelagerten Phasen des forensischen Prozesses bzgl. der Ursache eines Vorfalls und dessen Auswirkungen formuliert wurden, sicher und damit beweisbar zu bestätigen oder zu widerlegen. Die Korrelation aus einem Ereignis, das die Auswirkung einer zeitlich vorgelagerten Aktion ist, und eben dieser gesuchten Aktion gilt es durch die forensische Analyse zu bestimmen (Cooper et al. 2010). Die Erkenntnis hinsichtlich der Kausalität zweier Ereignisse kann dann in einer Sammlung aller relevanten Beweismittel, die im Zusammenhang mit einem Vorfall stehen, dokumentiert werden. Die so ermittelten Nachweise ermöglichen es, die Untersuchungsbereiche sukzessiv zu prüfen und abzuschließen. Durch dieses Vorgehen wird sichergestellt, dass die verfügbaren Datenbestände vollständig untersucht werden (Willer 2012, S. 293-339). Sofern die eingesetzten Untersuchungsmethoden und die daraus resultierenden Ergebnisse vollständig und nachvollziehbar dokumentiert werden, sodass das Analyseergebnis auf Basis der Originaldaten bei Bedarf reproduziert werden kann, sind die gesicherten Nachweise zivil- bzw. strafrechtlich verwertbar.

Damit kommt der Person bzw. der Gruppe von Personen, die mit der Datenanalyse betraut wurde, die Aufgabe zu, die in der Analysephase ermittelten und dokumentierten Nachweise inhaltlich, logisch und zeitlich zusammenzuführen. Sofern dies auf Basis der vorhandenen Nachweise nicht oder nur unvollständig möglich ist, müssen durch die Iteration vorgelagerter Phasen des forensischen Prozesses die dafür notwendigen Datenquellen bestimmt und gesichert werden. Infolgedessen ist es möglich, dass auch die forensische Datenanalyse mehrere Iterationen durchläuft, um neue Datenquellen und deren Nachweise in den bestehenden zeitlichen und logischen Verlauf der Ereignisse zu integrieren. In der abschließenden Dokumentation der Datenanalyse müssen alle Entwicklungsschritte und deren fachliche Auswirkung auf das Analyseergebnis dokumentiert werden. Im Kern der Datenanalyse stehen somit das Erkennen und Sichtbarmachen von Verbindungen und Abhängigkeiten innerhalb der Datenbestände, um die Ursachen für das Eintreten oder Ausbleiben eines Ereignisses zweifelsfrei bestimmen zu können. Auf dieser Grundlage ist es möglich, Vorfälle, die in einem informationstechnischen System oder durch ein solches entstanden sind, sowie deren Auslöser zu bestimmen (Dewald 2012, S. 38-43). Die Dokumentation der Erkenntnisse kann auch als Grundlage für reaktive oder präventive Maßnahmen dienen, die implementiert werden können, um die Ursache für das Auftreten oder Ausbleiben eines Ereignisses zu detektieren und zu verhindern (Beebe 2009, S. 17-36). Eine Implementierung der Maßnahmen erfolgt ggf. parallel zur Datenanalyse oder nachgelagert.

3.2.2 Vorgehen und Ziel

Das Vorgehen bei der forensischen Datenanalyse ist, verglichen mit anderen Bereichen der forensischen Wissenschaft, vielgestaltig, da die potenziellen Anwendungsfelder stets durch neue informationstechnische Systeme und Datentypen erweitert werden. Historisch bedeutet das, dass die Analysetätigkeit ein aufwendiger Prozess ist, bei dem Benutzer und Systemaktivitäten bestimmt, verarbeitet und analysiert werden müssen, um Erkenntnisse über die Art und den Ursprung eines Vorfalls zu ermitteln (Carrier und Spafford 2006). Die Zielsetzung, die in der Formulierung und Beantwortung von Ja-/Nein-Fragen liegt, ist dabei nicht immer und ggf. auch nicht vollständig auf Informationssysteme zu übertragen. Das liegt unter anderem daran, dass die Beschaffenheit eines Systems ggf. auf den Zustand eines einzelnen Attributs oder Objekts zurückzuführen ist. Ist dieser

Systemzustand nun Gegenstand eines Vorfalls, gilt es zu bestimmen, wann, wie und warum die Beschaffenheit des Systems verändert wurde (Geschonneck et al. 2014, S. 45-63). Im Rahmen der gesamten Vorfallsbearbeitung und damit auch der Datenanalyse gilt es dann, Indikatoren, die eine solche Bestimmung ermöglichen, zu identifizieren. Unterschiede, die bei der Bestimmung der Nachweise aufkommen, liegen häufig auch im Inhalt, in der Form und der Struktur des Beweismittels begründet. Aus ihnen leiten sich auch andere Anforderungen an die auszuwählenden forensischen Methoden ab (Carrier 2002, S. 1-3).

Digitale Aufnahmen wie Bild-, Video- oder Tondateien, die physisch existente Objekte und Sachverhalte abbilden, sind multimediazentriert und werden daher anders erfasst, verarbeitet und ausgewertet als Daten, die durch einen Netzwerk-Agenten oder einen Sensor in einer Industrieanlage erzeugt wurden. Bei der Erstellung von Daten, die ihren Ursprung innerhalb eines digitalen Systems haben und nur indirekt eine physisch existente Entität abbilden, wird häufig ein interpretierbares Format verwendet, das die Daten strukturiert (Labudde und Spranger 2017, S. 215-218). Infolgedessen ist es möglich, Kommunikations-, Verbindungs- und Protokolldaten weitestgehend automatisiert zu verarbeiten. Aufgrund einer fehlenden Syntax innerhalb unstrukturierter digitaler Aufzeichnungen, z. B. Bild- und Tonaufnahmen, setzt die automatisierte Verarbeitung dieser Dateiformate eine Vorverarbeitung voraus.

Nach der Erfassung und Verarbeitung der Datenbestände ist das Analyseziel zu definieren. Das Ziel der Analysetätigkeit ist kontextabhängig und kann daher variieren. So ist es möglich, dass Ermittler nach be- und entlastenden Beweisen innerhalb der verfügbaren Datenbestände suchen, um einen Tatverdacht zu erhärten und einen Beschuldigten zu identifizieren. Im Rahmen der Vorfallsbearbeitung in einem informationstechnischen System kann das Analyseziel jedoch auch darin bestehen, die Ursache eines technischen Fehlerzustandes und dessen Wirkung auf sicherheitsrelevante Ereignisse zu bestimmen. Die Dokumentation des Analyseergebnisses zu einem Sicherheitsvorfall kann auch dazu verwendet werden, Maßnahmen zu implementieren, mit denen Sicherheitsrisiken reduziert werden (Fowle und Schofield 2011). Wichtig sind hierbei Erkenntnisse, die es ermöglichen, die Anomalie oder den Vorfallsverlauf nachzuvollziehen und eine Folgenabschätzung, auch hinsichtlich weiterer Aktionen, durchzuführen.

Im Rahmen der Bearbeitung von sicherheitsrelevanten Ereignissen bedingt die Zielsetzung der IT-Forensik und somit der Datenanalyse die Untersuchung der folgenden Aspekte:

- Identifikation des Ereignisses, das den Vorfall hervorgerufen hat,
- Bestimmung der beteiligten Akteure und
- Erfassung, Verarbeitung, Analyse und Dokumentation der vorhandenen Nachweise.

Die Ermittlung dieser Aspekte kann im Kern durch die Beantwortung der folgenden grundlegenden W-Fragestellungen geleitet und abgeschlossen werden:

- Was ist passiert?
- Wo ist es passiert?
- Wie ist es passiert?
- Wann ist es passiert?
- Wer (Prozess, Nutzer, Aktion) hat es verursacht?
- Was kann getan werden, damit der Vorfall bzw. das Ereignis sich nicht wiederholt?

Das der Beantwortung der W-Fragen zugrunde liegende Konzept ermöglicht den Transfer von Erkenntnissen aus der Datenanalyse in eine allgemein verständliche Sprache, die in der Abschlussdokumentation zu einem Vorfalls Verwendung findet (Casey 2011, S. 255-282). Dieses Vorgehen ermöglicht die Zusammenfassung der Analyseerkenntnisse in einem Berichtsformat, durch das Vorfälle und Gesetzesverstöße entweder widerlegt oder aber nachvollziehbar bewiesen werden. Abbildung 4 beschreibt das Vorgehen während der Datenanalyse in Verbindung mit dem vorgelagerten Teilprozess, der Datenuntersuchung, und der im Anschluss an die Datenanalyse folgenden Dokumentation der Erkenntnisse in Form des Abschlussberichts.

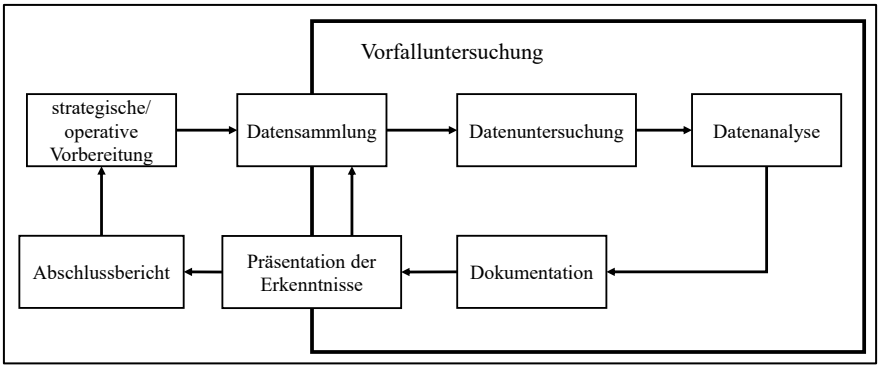

Abbildung 4: Vorgehen zur Vorfalluntersuchung (Eigene Darstellung)

Es ist auffällig, dass vorhandene Abhängigkeiten innerhalb des Gesamtprozesses dazu führen können, dass Prozessbestandteile mehrfach durchlaufen werden. Losgelöst von möglichen Iterationen einzelner Untersuchungsphasen ist während der Datenanalyse eine durchgehende prozessbegleitende Dokumentation zu erstellen.

3.2.3 Erfolgsbedingungen

Damit die Datenanalyse als Phase des forensischen Prozesses vor dem Hintergrund der zivil- und strafrechtlichen Verwertbarkeit der gesicherten Nachweise erfolgreich abgeschlossen werden kann, existieren Anforderungen an vor- und nachgelagerte Tätigkeiten, die Teil des Prozesses sind. Unabhängig vom Kontext des zu analysierenden Vorfalls können die folgenden Phasen als initiale Voraussetzung genannt werden (BSI 2011, S. 88-90):

- Datensammlung und Bergung:
  Digitale Aufzeichnungen, Geräte und datenverarbeitende Infrastrukturkomponenten, die im Kontext des Vorfalls für eine Untersuchung relevant sind, müssen identifiziert werden. Auf diese Weise wird sichergestellt, dass der Systemzustand bzw. die Integrität der Datenbestände nicht oder, wenn nicht anders möglich, nur in dem Umfang, der für die Sicherung notwendig ist, verändert wird. Die Datenbestände müssen also forensisch sicher extrahiert und aufbewahrt werden. Alle Aktionen, die im Rahmen der Sicherungsmaßnahmen unternommen werden, sind fortlaufend zu dokumentieren, sodass strukturelle und inhaltliche Änderungen der Daten nachvollzogen werden können.

- Untersuchung:
  Digitale Hilfsmittel und Werkzeuge, die im Rahmen der Datenuntersuchung eingesetzt werden, müssen bestimmte Kriterien in Bezug auf die Nicht-Veränderbarkeit der zu analysierenden Daten erfüllen. Daten, die im Teilprozess der Untersuchung als relevant identifiziert wurden, dürfen während der Extraktion nicht verändert werden. Jede auszuführende Transformation der Daten, die sich ggf. aufgrund möglicher Systembeschränkungen oder anderer Anforderungen ergibt, muss dokumentiert werden, sodass deren Ursache bekannt ist und die Transformation im Bedarfsfall vollständig wiederholt werden kann.

Datenbestände, die in diesen beiden Phasen gesichert und als relevant identifiziert werden, müssen unabhängig von den durchgeführten Transformationen und Tätigkeiten während der Datenanalyse in Form einer forensisch sicheren Kopie weiterhin bestehen. Daraus ergibt sich die Notwendigkeit, die Datenanalyse auf Grundlage eines forensischen Duplikats der gesicherten Datenbestände auszuführen. Die Datenanalyse ist losgelöst von Form, Ursprung und Struktur der Daten mit größtmöglicher Sorgfalt durchzuführen, sodass der Analyseprozess stets den Anforderungen an den forensischen Prozess genügt. Erkenntnisse, die während der Analyse gewonnen werden, können vorgelagerte Teilprozesse betreffen und dazu führen, dass Folgeaktionen ausgeführt werden müssen. Ein Beispiel hierfür ist die Notwendigkeit, Datenbestände zu sichern, die einen Zeitraum beschreiben, der zuvor nicht im Untersuchungsumfang definiert war. Diese Rückkopplung ist gleichzeitig eine Erfolgsbedingung, da das Analyseergebnis sonst unvollständig und damit nicht verwertbar wäre.

Es bestehen jedoch auch Erfolgsbedingungen, die ihren Ursprung in der strategischen und operationalen Vorbereitung haben. Ein Beispiel hierfür ist die Order of Volatility (dt. Reihenfolge der Flüchtigkeit). Diese wurde von der Internet Engineering Task Force (IETF) definiert und stellt eine Reihenfolge dar, in der Speicherinhalte beim Vorfinden eines Computers oder informationstechnischen Systems gesichert werden sollen. In dem entsprechenden Dokument, das unter RFC 3227 referenziert wird, ist die Notwendigkeit hinter der Rücksicherung von flüchtigen Speicherinhalten in den zu analysierenden Systemen beschrieben (Internet Engineering Task Force 2002). Sofern die Daten aus den Zielsystemen vollständig in dieser Reihenfolge gesichert werden können, steigt die Genauigkeit der ermittelten Nachweise. Umgekehrt bedeutet dies aber auch, dass, wenn

flüchtige Speicherinhalte nicht oder nur unvollständig gesichert werden können, das Analyseergebnis ggf. weniger genau und unvollständig ist. Sollten Datenbestände, die für die Analyse benötigt werden, aufgrund ihrer Flüchtigkeit nicht mehr rekonstruiert werden können, so ist dies entsprechend zu dokumentieren (Farmer und Venema 2005, S. 193-198). Die gleiche Dokumentationspflicht besteht für Datenbestände, bei denen im Rahmen der Analyse eine Manipulation durch Anti-Forensik-Maßnahmen und -Werkzeuge nachgewiesen werden kann. Darunter lassen sich Versuche und Maßnahmen gruppieren, die das Ziel haben, das Vorhandensein, die Menge oder die Qualität der Daten in einem Zielsystem negativ zu beeinflussen oder die Durchführung der Sicherungs- und Analysemaßnahmen zu erschweren bzw. unmöglich zu machen (Wundram et al. 2013, S. 83-97).

Trotz der Möglichkeit, dass Datenbestände, die im Rahmen des forensischen Prozesses verarbeitet und analysiert werden, unvollständig sind, gilt es, im Teilprozess der Datensicherung und -bergung bestehende Methoden und Mechanismen zur Datenerhebung konsistent zu nutzen. Anforderungen an Geräte, mit denen die Analyse ausgeführt werden soll, sind Bestandteil der strategischen Vorbereitung und variieren in Abhängigkeit von den zu analysierenden Systemlandschaften und deren Eigenschaften. Die folgenden Anforderungen sind allgemeingültig:

- Es sollten nur die Berechtigungen verwendet werden, die notwendig sind, um den Arbeitsschritt ausführen.
- Die Integrität des Untersuchungssystems sollte überprüfbar sein.
- Das Medium, auf dem die Datenbestände zu Analysezwecken gespeichert sind, sollte nur unter Verwendung eines Schreibschutzes genutzt werden.
- Das Untersuchungssystem und das zu untersuchende System sollten heterogen sein.

Die Beschränkung der Rechte in Bezug auf das Untersuchungssystem dient vor allem der Integrität dieses Systems. Mögliche Schäden, die durch die Ausführung von Schadsoftware auf dem zu untersuchenden System entstehen könnten, werden dadurch reduziert. Eine Möglichkeit, die Integrität des Untersuchungssystems sicherzustellen, liegt in der Verwendung eines Basis-Images, das über einen optischen Datenträger gestartet wird (Willer 2012, S. 49-94). Veränderungen an diesem Image können jederzeit festgestellt werden, was eine Überprüfung der Integrität des Untersuchungssystems bei

Bedarf ermöglicht und damit die Verwertbarkeit der Untersuchungsergebnisse sicherstellt.

3.3 Einsatzgebiete der forensischen Datenanalyse

Die IT-Forensik und damit auch die forensische Datenanalyse als elementarer Bestandteil des Untersuchungsprozesses sind Disziplinen der forensischen Wissenschaft. Schwerpunktmäßig werden hierbei digitale Beweise aus informationstechnischen Systemen verarbeitet und gespeichert. Ziele der forensischen Datenanalyse sind die Identifikation und die Analyse von Beweisen während eines Vorfalls oder nach diesem. Vor diesem Hintergrund erfolgt eine forensische Datenanalyse häufig anlassbezogen und ist somit stets Teil einer technischen Fehleranalyse, einer Untersuchung von zivil- und strafrechtlichen Verstößen, einer Prozessüberprüfung, von Revisionstätigkeiten oder anderen Sonderprüfungen. Aus diesen anlassbezogenen Untersuchungen lassen sich mögliche Einsatzgebiete der forensischen Datenanalyse ableiten (Carrier und Spafford 2006). Die im Kontext der forensischen Datenanalyse auszuführenden Maßnahmen und entsprechenden Möglichkeiten werden in den folgenden Unterkapiteln, je nach Einsatzgebiet, erläutert.

3.3.1 Zivil- und strafrechtliche Untersuchung

Werden Daten im Zuge eines zivil- oder strafrechtlichen Verstoßes analysiert, gibt es wesentliche Teilaspekte, die erfüllt werden sollten. Hierbei gilt es, Veränderungen im Laufzeitverhalten des Zielsystems zu identifizieren, sodass die Ursache für ein Ereignis bestimmt werden kann. Eine auf dem Zielsystem ausgeführte Handlung führt stets dazu, dass Prozesse ausgeführt oder beendet und durch Dateioperationen Elemente im Haupt- oder Festspeicher hinzugefügt, verändert oder entfernt werden. Diese Spuren ermöglichen es, einen Nachweis über eine ausgeführte Handlung und die damit verbundenen Änderungen am Systemzustand zu erstellen (Casey 2011, S. 35-47). Auf Grundlage dieser Nachweise ist es Ermittlern und Analysten möglich, Schlussfolgerungen über die Hintergründe der Ausführung und die ausführende Person zu ziehen. Diese Schlussfolgerungen basieren auf den Erfahrungen und dem Wissen des

Analysten oder Ermittlers zu den Funktionsprinzipien, die im betroffenen System Anwendung finden.

Zur Nachvollziehbarkeit des Vorfalls werden die Nachweise häufig gemäß ihrer zeitlichen Abfolge dargestellt, sodass es möglich ist, die eigentliche Handlung und nachgelagert auch die durch sie bedingten Veränderungen am Zielsystem nachzuvollziehen. Eine solche Zeitachse kann mit Metadaten aus anderen beteiligten Systemen und Verbindungs- und Kommunikationsdaten angereichert werden, sodass sich daraus die Ursächlichkeit eines Vorfalls und einer vorgelagerten Aktion bestimmen lässt. Eine alternative Möglichkeit, den Vorfallsverlauf darzustellen, ist die Why-Because-Analyse (Sanders 2012). Aus dieser lässt sich eine Grafik ableiten, aus der hervorgeht, welches Ereignis den zu untersuchenden Vorfall ausgelöst hat. Unter Berücksichtigung dieser Information ist es prinzipiell möglich, nachzuweisen, dass ein Vorfall bzw. granularer auch ein einzelnes Ereignis stattgefunden hat. Der eindeutige Nachweis, dass die beschuldigte natürliche Person auch für die Ausführung verantwortlich ist, lässt sich auf Basis dieser Information jedoch nicht immer zweifelsfrei führen. Dieser Nachweis kann häufig nur unter Berücksichtigung weiterer Informationsquellen erfolgen. Durch die Kombination von Nachweisen, die im Rahmen der forensischen Datenanalyse ermittelt wurden, und Aussagen oder Beweisen, die für oder gegen ein mögliches Alibi einer beschuldigten Person sprechen, ist es möglich, den Täter, das Tatmittel und die vom Vorfall betroffenen Personen oder Systeme zu bestimmen. Die Visualisierung in Form eines Zeitstrahls, auf dem die Ereignisse des Vorfalls abgebildet werden, ermöglicht auch außenstehenden Gutachtern und beteiligten Personen, die einzelnen Geschehnisse des Vorfalls nachzuvollziehen und diese einzuordnen.

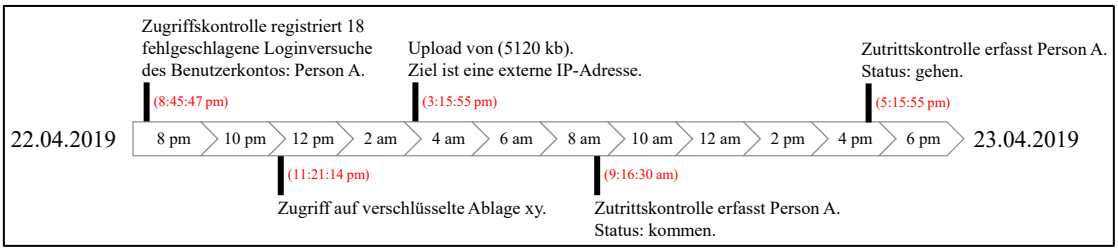

Abbildung 5: Zeitstrahl zur Darstellung eines strafrechtlichen Verstoßes (Eigene Darstellung)

Das in Abbildung 5 dargestellte Beispiel stellt einen Zeitstrahl zur Darstellung eines strafrechtlichen Verstoßes in einem informationstechnischen System dar.

3.3.2 Technische Fehleranalyse

Die Bewertung eines sicherheitsrelevanten Vorfalls in einem informationstechnischen System erfolgt häufig durch ein CERT. Handelt es sich bei dem Vorfall um eine gezielte, bewusst ausgeführte Handlung mit der Absicht, die Systemfunktion einzuschränken oder Daten, die in dem System gespeichert und verwaltet werden, zu verändern oder abzugreifen, kann dieses Vergehen zivil- oder strafrechtlich verfolgt werden.

Liegt die Ursache eines Vorfalls in einem System, wobei ein Fremdverschulden sicher ausgeschlossen werden kann, ist es möglich, mit den Mitteln der IT-Forensik Nachweise zur Beschreibung der Anomalie zu identifizieren, zu sichern und zu analysieren. So können mögliche Angriffsszenarien und Aktionen, die der Vorbereitung eines solchen Angriffs dienen, ausgeschlossen werden. Beispielhaft ist hier das aktive und passive Footprinting zu nennen, das die erste Phase eines Hackerangriffs ist (Mihailowitsch 2011). Hierzu kann ein Angreifer öffentlich zugängliche Quellen oder illegale Anwendungen nutzen, um Informationen über ein Zielsystem zu beschaffen. Im Anschluss daran kann das System auf Basis der gesicherten Datenbestände auf Interdependenzen mit anderen Systemen und fehlerhafte oder unvollständige Konfigurationsparameter geprüft werden.

Die erfassten Datenbestände ermöglichen eine deskriptive Darstellung des Systemverhaltens, sodass es ab dem Zeitpunkt des Vorfalls möglich ist, ausgeführte Aktionen und deren Auswirkungen auf das System zu bestimmen. Aufbauend auf diesen Erkenntnissen können Anpassungen an den betroffenen Systemen vorgenommen werden. Aufgrund der Anforderungen an die forensische Datenanalyse in und nach den Phasen der Sicherung, Verarbeitung und Analyse ist es möglich, die Funktion der Anpassung mit einem A/B-Test zu überprüfen. Die in Tabelle 1 dargestellte Übersicht ermöglicht es, im Rahmen der technischen Fehleranalyse Systeme und Anwendungskomponenten zu bestimmen, mit deren Datenströmen die Ursachen für Fehlerzustände in komplexen Informationssystemen bestimmt werden können.

Informationsquelle	Art der Information
Firewall-Verkehrsprotokoll	Informationen über ein- und ausgehenden Netzwerkverkehr sowie erlaubte und unerlaubte Netzwerkverbindungen
Mail-Sicherheitsgateway	Informationen über den eingehenden E-Mail-Verkehr, unter anderem die Anzahl der schadcodebehafteten und blockierten (Spam-)Nachrichten
Intrusion-Prevention-/ Intrusion-Detection-Systeme	Warnmeldungen über verdächtigen und blockierten Netzwerkverkehr
Systemüberwachung	Informationen über die Systembetriebszeit, den Systembetriebszustand, die Ping-Latenz und die Netzwerk-, Hauptspeicher- und Prozessorauslastung
Schwachstellen-Scan	Erkennung von Systemschwachstellen, Deltas in der Schwachstellenbehebung, Änderungen an offenen/geschlossenen Schnittstellen
Logsammler	Informationen aus Syslog, Windows Event Log, Simple Network Management Protocol (SNMP)
Dateiüberwachung	Benachrichtigung über Änderungen an sensiblen Daten oder Konfigurationsänderungen

Tabelle 1: Informationsquellen im Rahmen der technischen Fehleranalyse

Die nach den Maßstäben des forensischen Prozesses ausgeführte Datenanalyse ist stets ziel- und lösungsorientiert, sodass es allen Akteuren, die an der Suche und Behebung eines Fehlerzustands beteiligt sind, möglich ist, das ursächliche Verhalten zu bestimmen und reaktive Maßnahmen einzuleiten.

3.4 Herausforderungen vor und während der Datenanalyse

Die Notwendigkeit, auch Massendaten im Rahmen des forensischen Untersuchungsprozesses zu sichern, zu verarbeiten und zu analysieren, führt zu praktischen Herausforderungen vor und während der Ausführung verschiedener Prozessphasen. Die Verarbeitung der Daten aus dem Quellsystem, die zum einen in der physischen Verarbeitung der Daten, in Form der effizienten Speicherung des forensischen Abbilds, und zum anderen in der inhaltlichen Verarbeitung der Dateninhalte besteht, ist ein Beispiel für eine solche Herausforderung. Bestimmbare Grenzen bzgl. des verarbeitbaren Datenvolumens gibt es grundsätzlich nicht. Zu beachten ist aber, dass die Komplexität der Analyse, neben Faktoren wie der Struktur und dem Inhalt der Daten, auch wesentlich vom zu analysierenden Datenvolumen abhängt. Maßnahmen, mit denen das Ziel verfolgt wird, relevante Inhalte innerhalb einer Sammlung von Massendaten zu identifizieren, lassen sich nur bedingt auf Datenquellen anwenden, deren Inhalt und Form unstrukturiert ist. Beispiele hierfür sind Bild- und Tonaufnahmen sowie E-Mail-Postfächer. Das Risiko, dass Daten bereits vor der Erhebung als nicht relevant markiert werden, ist damit allgegenwärtig und führt ggf. dazu, dass Nachweise mit einem hohen Beweiswert bereits während der Datensicherung nicht berücksichtigt werden (Casey 2011, S. 465-510). In diesen Fällen ist es möglich, Experten, die über ein hohes Maß an Fachwissen bzgl. der zu analysierenden Domäne verfügen, oder bestehende Leitfäden des BSI bzw. der International Organization for Standardization zu konsultieren, um sicherzustellen, dass die für die Analyse notwendigen Daten vollständig gesichert werden. Eine weitere Herausforderung besteht darin, dass es informationstechnische Systeme und Anwendungen gibt, bei denen kein exaktes Abbild des Systems erstellt werden kann, ohne dass dabei das Zielsystem selbst, wenn auch nur marginal, verändert wird. Sollte es während der Datensicherung dazu kommen, dass Maßnahmen und Werkzeuge eingesetzt werden müssen, die den Zustand des Systems verändern, so sind diese Änderungen der Datenbasis entsprechend zu dokumentieren. Dies gilt unabhängig von der allgemeinen Dokumentationspflicht zur Wahrung der zivil- und strafrechtlichen Verwertbarkeit der zu ermittelnden Nachweise.

Eine weitere zentrale Herausforderung bildet die Order of Volatility. Werden Daten entgegen dieser Reihenfolge identifiziert und archiviert, führt das dazu, dass Nachweise ggf. durch die Sicherungsaktion unwiederbringlich überschrieben oder verändert werden.

Besonders betroffen sind hiervon die Datenbestände in den Registern, in Cache-Speichern, der Hauptspeicherinhalt, aktive Prozesse und Netzwerkverbindungen (Parasram 2017, S. 80-96). Sind die zu analysierenden Daten aus den Quellsystemen extrahiert und gesichert worden, besteht der Anspruch, die Datenbestände effektiv auszuwerten. Hierzu gibt es vielfältige Werkzeuge, mit denen Analysten oder Ermittler die Datenbestände effektiv auswerten können. Da eine inhaltliche Analyse von Bild- und Tondaten oder großen semistrukturierten Textinhalten mit einem hohen manuellen Aufwand verbunden ist und nur bedingt durch andere Prozessbeteiligte nachvollzogen werden kann, gibt es Methoden, mit denen die zu analysierenden Datenbestände visuell aufbereitet werden. Auf dieser Grundlage ist es allen Prozessbeteiligten möglich, zeitliche und inhaltliche Zusammenhänge innerhalb der Datenbestände wahrzunehmen. Gleichzeitig ermöglichen diese Methoden eine vereinfachte Bewertung und Analyse von gleichförmigen Massendaten, indem die einzelnen Elemente in aggregierter Form dargestellt werden. Herausforderungen und Einschränkungen innerhalb des forensischen Prozesses, die sich aus dem zunehmenden Datenvolumen ergeben, können mithilfe dieser Methoden reduziert werden.

4 Die Daten- und Informationsvisualisierung

Daten, die in Informationssystemen erzeugt, verarbeitet und gespeichert werden, sind in ihrer Struktur heterogen. Sensordaten, die in einer Industrieanlage erzeugt werden, unterscheiden sich wiederum von Datenbeständen, die durch Finanzmarkttransaktionen entstehen. Die Struktur und Komplexität der anfallenden Daten ergibt sich unabhängig vom Quellsystem aus dem Inhalt, dem Format und dem Volumen der Quelldaten. Um Datenbestände für den Menschen sichtbar und wahrnehmbar zu gestalten, haben sich bestimmte Gestaltungsformate etabliert (Scheuermann 2010). Ein Beispiel ist die Darstellung und Beschreibung von Daten in Form von Text, Zahlen und Grafiken. So ist die Analyse von Massendaten theoretisch ohne Verwendung von Hilfsmitteln möglich. Dieses Vorgehen ist in der Durchführung jedoch ressourcenintensiv und damit praktisch irrelevant. Die Analyse und Anomalieerkennung bei ein- und ausgehenden Netzwerkverbindungen, die von einer Firewall aufgezeichnet werden, können als Beispiel dafür angeführt werden: Bei drei-, vier- oder fünfstelligen Ereignisraten pro Sekunde wird deutlich, warum es nicht möglich ist, diese, in einem technischen Format erzeugten und zwischengespeicherten, Verbindungsversuche ohne Hilfsmittel zu bewerten (Card et al. 1999, S. 625-637).

Ein erprobtes Hilfsmittel zur Behebung der Problemstellung ist die Informationsvisualisierung. Der Nutzen der Informationsvisualisierung liegt darin, einen Datensatz vollständig zu erfassen, ihn zu verarbeiten und daraus eine exakte visuelle Darstellung der Daten zu generieren (Huang et al. op. 2005). Das Ziel dieser Visualisierung ist es, dass der Benutzer die im Datensatz enthaltene Information selbstständig erfassen kann. Hierzu bedarf es eines vollständigen Überblicks über die Informationen des Datensatzes, sodass der Betrachter die Möglichkeit hat, Muster und Beziehungen innerhalb der zugrunde liegenden Daten zu erkennen. Um aus diesem Überblick weitere Detailinformationen abzuleiten, wird über eine zu definierende Auswahl innerhalb des bestehenden Datensatzes ein neuer, dem Zielkorridor entsprechender Datensatz generiert. Dieser bietet dann die Grundlage für eine detailliertere Bewertung der Bestandsdaten.

Die Zielstellung dessen ist stets die verständliche Vermittlung von Informationen an mögliche Interessensgruppen. Der Grad der Zielerreichung ist jedoch abhängig von der Datengrundlage. Informationen und Daten werden daher so angereichert, dass sie von der Zielgruppe interpretiert werden können. Geografische Informationen und datums- und zeitbasierte Daten eignen sich bspw. für die Erstellung einer geografisch-historischen Übersicht, die es dem Betrachter ermöglicht, einen räumlichen oder zeitlichen Bezug zum beschriebenen Ereignis herzustellen. Ein Beispiel hierfür ist die Beschreibung von numerischen Daten wie der Einwohnerzahl, dem Bruttoinlandsprodukt oder der Bevölkerungsdichte je Kontinent, Land, Provinz und Stadt durch das Einfärben des Plan- oder Karteninhalts. Visualisierungen unterstützen kognitive Prozesse des Menschen insofern, als dass die Menge der zur Verfügung stehenden Informationen reduziert wird und damit sowohl die Erfassung als auch die Verarbeitung dieser Informationen ermöglicht werden.

4.1 Hintergrund und Entstehungsgeschichte

Zur Vermittlung von Informationen, die auf einzelnen Datensätzen oder einer beliebigen Menge an Daten beruhen, war es historisch und ist es noch heute notwendig, diese Datensätze so exakt wie möglich wiederzugeben. Dem Informationsempfänger sollte es möglich sein, einen uneingeschränkten Überblick über die Information zu erhalten und wenn nötig die zugrunde liegenden Daten, aus denen die Information abgeleitet wurde, einzusehen. Zur Erstellung einer Übersicht, die es den Empfängern ermöglicht, Zusammenhänge, Muster und Abweichungen innerhalb der darzustellenden Daten zu erkennen, wurden im 18. Jahrhundert erstmals verschiedene Diagrammtypen eingesetzt (Playfair et al. 2005, S. 1-23). Diese Form der Darstellung war zunächst auf Zeitreihen beschränkt, die jedoch noch im gleichen Jahrhundert durch weitere Diagrammtypen, wie Linien- und Balkendiagramme, ergänzt wurden. Diese frühen Maßnahmen zur Visualisierung der Daten hatten die Gemeinsamkeit, dass sie eine Beziehung zwischen den zugrunde liegenden Daten sichtbar machten. Ein Beispiel hierfür ist das in Abbildung 6 dargestellte Balkendiagramm, das die Importe und Exporte Schottlands im Jahr 1781 abbildet.

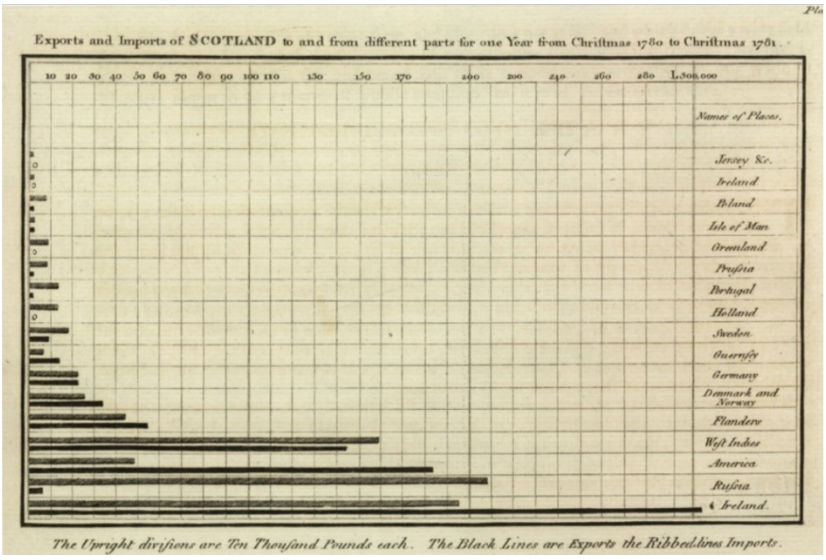

Abbildung 6: Balkendiagramm zu den Importen und Exporten Schottlands aus dem Jahr 1781 (Playfair 1786)

Diese Darstellungsform ermöglichte es erstmals, umfangreiche Datenbestände und deren Bedeutung, wie hier den Austausch von Waren, Dienstleistungen und Kapital über internationale Grenzen hinweg, allgemein verständlich und vergleichbar darzustellen. Aufgrund der Beliebtheit solcher Schaubilder wurden neue Darstellungsformen entwickelt. Mit der Einführung von Bildstatistiken wurden dann auch Mengengerüste und deren Veränderung über den, vom Ersteller der Grafik gewählten, Darstellungszeitraum sichtbar (Brinton 1914, S. 20-52). Bei der in Abbildung 7 dargestellten Grafik handelt es sich um eine Bildstatistik. In dieser werden Güterverkehrsleistungen, die in den Vereinigten Staaten auf dem Schienenweg erbracht wurden, in Tonnen und jeweils für die Jahre 1899 und 1911 dargestellt

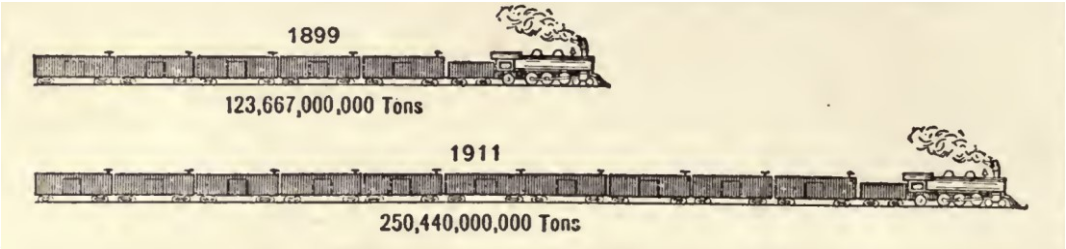

Abbildung 7: Bildstatistik zur Gesamtzahl der Güterverkehrsleistungen mittels Eisenbahnen aus dem Jahr 1914 (Brinton 1914, S. 40)

Die Informationsvisualisierung ist historisch gesehen also ein Hilfsmittel, mit dem Rohdaten und die mitunter komplexen Inhalte, die durch sie beschrieben werden, in

einem verständlichen Format dargestellt werden, sodass die Inhalte verschiedenen Interessensgruppen vermittelt werden können (Brinton 1914, S. 321-343). William Playfair, Charles Joseph Minard, Otto Neurath und Nigel Holmes prägten diese Form der Informationsdarstellung insofern, als dass die Darstellungsstruktur bis heute bekannt ist und aktiv eingesetzt wird. Mit dem Einzug der Datenverarbeitung in technischen Anlagen wuchs die Menge an darzustellenden Daten und durch den Einsatz von grafischen Benutzeroberflächen war es möglich, neue Diagramme und Darstellungsformen zu entwickeln. Ben Shneiderman veröffentlichte im Jahr 1996 hierzu einen Beitrag mit dem Titel ‚Visual Information Seeking Mantra'. In diesem definierte er das folgende Vorgehen: overview first, zoom and filter, then details-on-demand (Shneiderman 1996, S. 336-343). Diese Beschreibung legt nach Shneiderman das allgemeine Vorgehen eines Benutzers bei einer Visualisierungsanwendung fest.

Heute werden Visualisierungen und Informationsgrafiken dafür verwendet, komplexe Datensätze in visuelle Modelle und Darstellungen zu überführen. Diese ermöglichen verschiedenen Interessensgruppen einen systematischen Zugang zu den Inhalten des Datensatzes, indem Zusammenhänge und Abläufe, deren Entstehung, eine räumliche und zeitliche Einordnung oder ein sachlicher Bezug dargestellt werden.

4.2 Grundlagen der Daten- und Informationsvisualisierung

Die visuelle Darstellung von Daten ist ein effektives Hilfsmittel, das dazu genutzt werden kann, Informationen zielgruppenorientiert oder allgemein verständlich aufzubereiten. Im Rahmen dieses Kapitels werden daher zentrale Aspekte der Daten- und Informationsvisualisierung erläutert. Der Zweck und die Funktion der Begriffe ‚Information' und ‚Informationsvisualisierung' werden hierzu bestimmt und abgegrenzt. Auf dieser Grundlage ist es möglich, den Nutzen der Daten- und Informationsvisualisierung zu benennen und zu erläutern. Die Voraussetzungen und Maßnahmen zur Erstellung effektiver Visualisierungen bilden dann den Übergang zur Beschreibung lebensnaher Einsatzmöglichkeiten der Informationsvisualisierung und deren Implementierung.

4.2.1 Allgemeine Definition

Der Begriff ‚Information' beschreibt eine beliebige Datenmenge, der aufgrund einer Eigenschaft eine Bedeutung zugeordnet werden kann. Wird eine Information visualisiert, heißt das, dass Konzepte, Methoden und Werkzeuge zur visuellen Darstellung von Informationen aus Datenbanken, digitalen Bibliotheken oder anderen Dokumentensammlungen angewendet werden. Diese visuelle Darstellung ermöglicht es dem Betrachter, eine Beziehung zwischen Daten, also den Bestandteilen der Information, wahrzunehmen. Sprachlich oder inhaltlich nur schwer formulierbare Zusammenhänge werden über den Prozess der Visualisierung inhaltlich verständlich aufbereitet. Die Möglichkeiten der Informationsvisualisierung sind grundsätzlich interdisziplinär und damit über verschiedene Domänen hinweg einsetzbar. Es ist möglich, Prozessabläufe, Objektrelationen und spezifische Informationen aus Sammelbeständen zu identifizieren und diese im Kontext mit anderen Informationen aus unterschiedlichen Sichtweisen zu beschreiben. Der amerikanische Informatiker Shneiderman hat dieses Vorgehen teilweise in seinem ‚Visual Information Seeking Mantra' beschrieben: „overview first, zoom and filter, then details on demand" (Shneiderman 1996, S. 336-343). In Bezug auf dieses Paradigma, das ausgehend von einem Überblick die Möglichkeit fordert, die Daten zu selektieren, um sie bei Bedarf detailliert einzusehen, ist jegliche Form der Visualisierung effektiv, sofern sie denn eine mehrstufige Informationsbeschaffung aufgrund von Bestandsdaten ermöglicht. Shneidermans Vorgehen basiert auf der Annahme, dass das Ziel bei der Informationsbeschaffung im Regelfall darin besteht, einen Datenpunkt oder eine Gruppe von Datenpunkten zu identifizieren. Die Visualisierung wird daher als Hilfsmittel zur Bestimmung dieser Teilmenge beschrieben (Card et al. 1999, S. 579-624). Während der Bestimmung dieser Teilmenge dient sie als Möglichkeit zur Erkennung von Mustern und Abweichungen innerhalb des Datensatzes.

4.2.2 Nutzen der Datenvisualisierung

Grafiken, Diagramme oder Animationen, in denen Daten dargestellt werden, lassen sich unter dem Begriff ‚Visualisierung' zusammenfassen. Werden Daten visualisiert, unterstützt dies die kognitive Wahrnehmung des Menschen, indem die Visualisierung als Zwischenspeicher fungiert und damit das Arbeitsfeld erweitert wird (Hutchison et al. 2008, S. 1-18). Hieraus ergeben sich Wahrnehmungsvorteile für den Betrachter, da

grafisch aufbereitete Informationen parallel interpretiert werden können. Textuell aufbereitete Informationen werden im Vergleich hierzu sequenziell interpretiert (Hendee und Wells 1997, S. 27-33). Grafisch aufbereitete Informationen können somit schneller interpretiert werden können als textuell dargestellte. Die Grundlage hierfür bildet die Gestalttheorie, die die Wahrnehmung von Objekten und Mustern beschreibt und sich aus den folgenden Gestaltgesetzen ergibt:

- Figur-Grund: Es gibt Merkmale, die bei der Wahrnehmung hervorstehen und Merkmale, die in den Hintergrund treten. Es ist nicht möglich, beide Merkmalstypen gleichzeitig wahrzunehmen.
- Nähe: Objekte, die räumlich nahe beieinander sind, werden als Gruppe wahrgenommen.
- Gleichheit: Ähnliche Objekte werden als Gruppe wahrgenommen.
- Kontinuität: Fließende, ununterbrochene Formen werden eher wahrgenommen als diskontinuierliche Formen.
- Geschlossenheit: Unvollständige Objekte werden durch die Wahrnehmung geschlossen oder vervollständigt.
- Symmetrie: Symmetrische Objekte werden als zentriert und als Ganzes wahrgenommen.
- Erfahrung: Komplexe oder unbekannte Erscheinungen werden unter Zuhilfenahme von bereits bekannten Objekten interpretiert.

Die Gestaltung effektiver Visualisierungen kann unter Berücksichtigung dieser Prinzipien erfolgen (Zimbardo et al. 2008, S. 143-146).

Der gesteigerte Nutzen einer Visualisierung, in Bezug auf die Wahrnehmung der zu vermittelnden Inhalte, wird häufig erst unter Berücksichtigung des Einsatzzwecks deutlich. Visualisierungen werden eingesetzt, wenn Zusammenhänge oder Prozessabläufe, Zahlenwerte und deren Verhältnis oder Ereignisse und deren Auswirkungen räumlich beschrieben werden sollen. Die visuelle Darstellung bietet grundsätzlich die Möglichkeit, große Datenmengen und deren Inhalte vergleichbar zu machen, da Veränderungen und deren Auswirkungen auch historisiert beschrieben werden können. Auf dieser Grundlage ist es möglich, die Auswirkungen eines Ereignisses in Relation zu anderen Ereignissen und deren Verlauf zu setzen und dadurch Muster und Abhängigkeiten zu identifizieren. Die Möglichkeiten der Visualisierung sind jedoch nicht

nur darauf beschränkt, komplexe Datensätze zugänglicher zu machen, ihr Nutzen liegt auch darin, relativ einfache Datensätze aufzubereiten, sodass deren Inhalt attraktiver dargestellt und einfacher zu vermitteln ist. Damit bildet die Visualisierung eine effektive Möglichkeit, Betrachter in Themenfelder einzuführen, zu denen sie zuvor über wenig bis kein Wissen verfügten. Mögliche Hintergründe zum Aufbau und Verständnis des Inhaltes können ggf. schneller vermittelt werden, als wenn die gleichen Inhalte rein textuell, über einen Aufsatz zum Thema, vermittelt werden.

Ein historisches Beispiel, das die Möglichkeiten und den Wert der Informationsvisualisierung bei der Analyse von Informationen zeigt, ist die Arbeit des Epidemiologen Dr. John Snow, dargestellt in Abbildung 8 (Barton 2018).

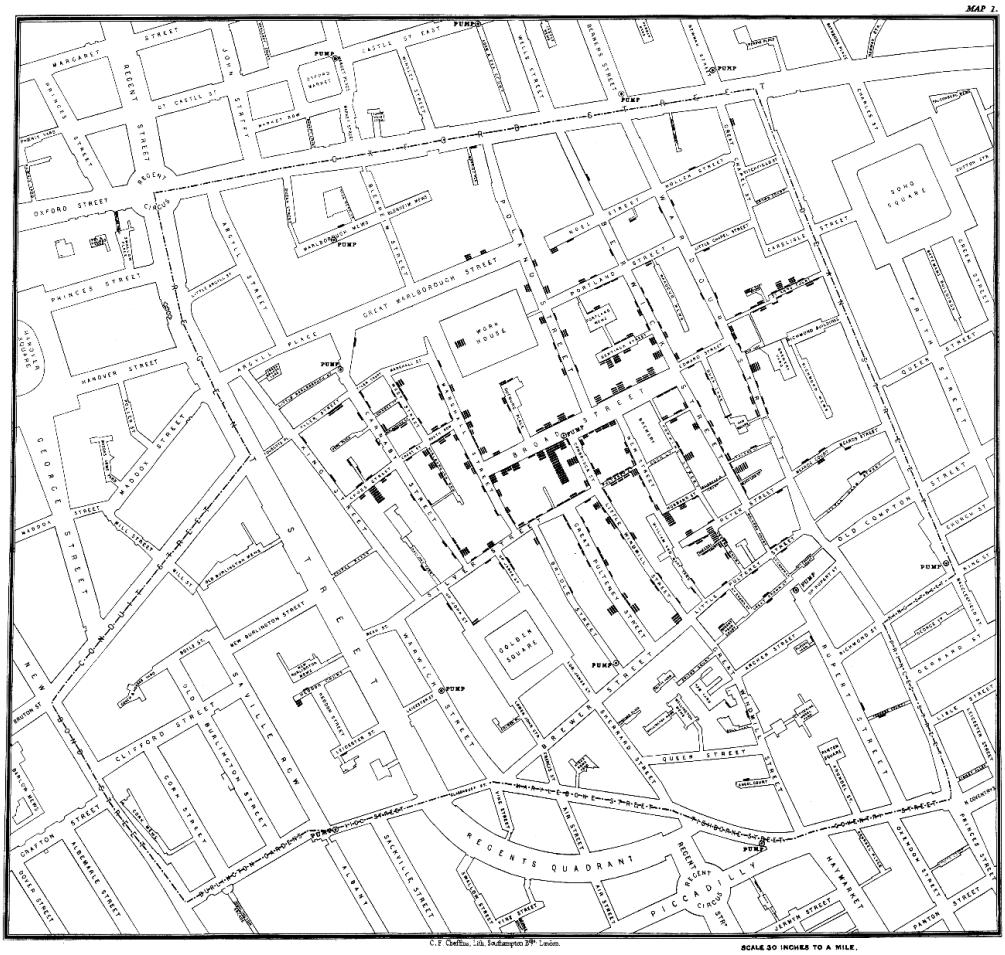

Abbildung 8: Informationsvisualisierung am Beispiel bestätigter Cholerainfektionen (Barton 2018)

Durch seine Darstellung von bestätigten Cholerainfektionen im Jahr 1842 auf einer Karte des Stadtzentrum Londons wurde ein Muster sichtbar. Auf Grundlage dieses Musters war es möglich, den Ursprung der Cholerabakterien auf verunreinigtes Trinkwasser zurückzuführen und die betroffenen Quellen ausfindig zu machen. Die Visualisierung bietet folglich zwar kein Mittel zur Heilung des Krankheitsbildes, jedoch bietet sie eine Entscheidungsgrundlage für die Einleitung reaktiver Maßnahmen, die es wiederum ermöglichen, die Verbreitung einzudämmen. Es sind dieser Aspekt und die daraus resultierenden Folgen, die die Visualisierung und ihren Nutzen maßgeblich beschreiben und dazu führen, dass über alternative Darstellungen und daraus hervorgehende Sichtweisen neue Erkenntnisse aus digitalen Datenbeständen gewonnen werden können.

4.2.3 Erstellung effektiver Visualisierungen

Die Erstellung einer wirksamen Visualisierung, die es dem Betrachter ermöglicht, die dargestellten Inhalte vollständig zu erfassen, hat Shneiderman im ‚Visual Information Seeking Mantra' beschrieben (Shneiderman 1996, S. 336-343). In diesem wird das Verhalten eines Betrachters erläutert und es werden Voraussetzungen genannt, die durch den Ersteller der interaktiven Visualisierung erfüllt werden sollten. Hierbei handelt es sich um die folgenden sieben Teilschritte:

1. Übersicht:
   Die Visualisierung sollte einen Überblick über alle verfügbaren Daten ermöglichen. Ist es aufgrund der Quantität der Datensätze nicht möglich, sie gesammelt dazustellen, können die Daten gemäß einer Eigenschaft vorselektiert werden, sodass dann diese Gruppierung angezeigt werden kann.
2. Zoom:
   Einzelne Elemente oder Gruppen von Elementen sollten innerhalb der Gesamtdaten sichtbar gemacht werden können. Eine Zoomfunktion ermöglicht es, die ausgewählten Elemente anzuzeigen, ohne dabei Hintergrundinformationen und den Kontext der Darstellung auszublenden.
3. Filter:
   Es sollte dem Benutzer der Visualisierung möglich sein, Elemente, die für ihn von Interesse sind, auszuwählen und bei Bedarf andere Elemente ein- oder auszublenden. Umgekehrt sollten Elemente, die für den Benutzer irrelevant sind,

bei Bedarf ausgeblendet werden können. Ziel dieses Filters ist es, die Informationen zu selektieren, sodass nur die im Kontext der Analyse notwendigen Informationen dargestellt werden.

4. Details auf Anfrage:
   Wird ein einzelnes Element innerhalb der Visualisierung ausgewählt, sollte es möglich sein, die dem Wert zugrunde liegenden Rohdaten einzusehen. Dies gilt insbesondere für aggregierte Werte und Ereignisse, die z. B. nach Minuten, Stunden oder Tagen gruppiert wurden.
5. Zugehörigkeit:
   Es sollte dem Betrachter möglich sein, Beziehungen zwischen verschiedenen Elementen zu erkennen und zu markieren. Die Visualisierung sollte hierzu über die Möglichkeit der Elementbeschriftung verfügen, mit der ähnliche Elemente auf Grundlage gleicher Attribute identifiziert werden können.
6. Verlauf:
   Werden durch den Benutzer der Visualisierung Aktionen ausgeführt, die dazu führen, dass die Daten oder deren Darstellung verändert wird, so sollten die ausgeführten Aktionen und deren Auswirkungen erfasst und gespeichert werden. Auf Grundlage dieser historisierten Übersicht sollte es dem Nutzer möglich sein, ausgeführte Aktionen zu wiederholen oder rückgängig zu machen. Die Gesamtdokumentation ermöglicht es, auch die Erstellung der Visualisierung im Bedarfsfall zu wiederholen und bei wiederkehrenden Untersuchungen zu optimieren.
7. Extraktion:
   Dem Benutzer sollte es möglich sein, die von ihm erstellte oder angepasste Visualisierung zu speichern und bei Bedarf zu exportieren.

Unter Berücksichtigung dieser Anforderungen lässt sich die folgende Aussage über die Erstellung und Nutzung interaktiver Grafiken ableiten: Eine Visualisierung kann wirksam eingesetzt werden, wenn die gewählte Art der grafischen Darstellung es dem Benutzer ermöglicht, die im Datensatz enthaltenen Informationen wahrzunehmen und die Darstellung gemäß seinen Anforderungen anzupassen (Meirelles 2013, S. 6-15).

4.3 Einsatzgebiete der Daten- und Informationsvisualisierung

Die Bereiche, in denen Daten und Informationen visualisiert werden, können sind vielfältig. Eine Bestimmung und Unterscheidung der Visualisierungsmöglichkeiten nach deren Einsatzgebiet erfolgt häufig durch die Domäne und den Anwendungskontext, in dem die Daten dargestellt werden sollen. Weit verbreitet sind Visualisierungen im Kontext von bildgebenden Verfahren in der medizinischen Diagnostik, bei der Produktdarstellung und bei wissenschaftlichen Untersuchungen. Da der Begriff ‚Informationsvisualisierung' alle visuellen Darstellungen von Daten umfasst, ist er nicht auf einzelne Einsatzgebiete oder Domänen beschränkt, sondern kann beliebig erweitert werden (Card und Mackinlay 1997, S. 92-99). Anwendungsgebiete im Kontext von informationstechnischen Systemen sind hierbei von besonderer Bedeutung. Aufgrund des großen Datenvolumens, das während der Implementierung und des Betriebs dieser Systeme erzeugt, verarbeitet und gespeichert wird, sind sie für die Entwicklung neuer und die Optimierung bestehender Visualisierungsformen prädestiniert. Werden Darstellungs- oder Verarbeitungsgrenzen während der Visualisierung von semi- oder unstrukturierten Massendaten identifiziert, so hat dies einen Einfluss auf Forschungsbemühungen, da sich die Forscher dieser Herausforderungen annehmen. Sofern bestehende Problemstellungen durch die Entwicklung und Einführung neuer Methoden gelöst werden können, können die mitunter positiven Effekte in Bezug auf die Robustheit und die Fehlertoleranz der Darstellungsformen auch in anderen Anwendungsfeldern genutzt werden.

Ein verbreitetes Anwendungsgebiet der Informationsvisualisierung ist die explorative Analyse von großen Datensätzen. Die Visualisierung eines Datensatzes kann jedoch auch anlassbezogen erfolgen, wobei die Zielsetzung häufig in der schnellen Eingrenzung des Suchraums besteht, um z. B. vor dem Hintergrund einer Fragestellung auskunftsfähig zu sein. Die Visualisierung eines Datensatzes bietet in ihrer initialen Form keine Antworten auf spezifische Fragestellungen, sondern vielmehr die Grundlage für die Formulierung präziser ereignisbezogener Anfragen. Die allgemeingültige und objektive Bewertung der Wirksamkeit von Visualisierungen und deren Einfluss auf die Beantwortung spezifischer Anfragen, die zu Erkenntnisgewinnen führen, ist indes nicht möglich, da die zu erzielenden Ergebnisse nur bedingt quantifizierbar sind (Ahn et al. 2011, S. 1-8). Im Rahmen der Vorfallsbearbeitung könnte eine Bewertung der Wirksamkeit über die

Zeitspanne, die zwischen der Erstellung der Frage und deren Beantwortung verstrichen ist, erfolgen. Weitere Vergleichsmöglichkeiten bestehen in Bezug auf die Genauigkeit der gewählten Darstellungsform, die Bedienbarkeit und den Aufwand, der durch den Anwender der Visualisierung bzw. Visualisierungsmethode erbracht werden muss, um den Datensatz zu visualisieren und Erkenntnisse abzuleiten.

In den folgenden Unterkapiteln soll daher ein Ausblick auf die Anwendungsgebiete im Kontext von informationstechnischen Systemen, insbesondere im Hinblick auf die Verarbeitung und Darstellung von sicherheitsrelevanten Massendaten, gegeben werden. Hierzu wurden die Anwendungsgebiete Electronic Discovery (eDiscovery) und Log Management ausgewählt.

4.3.1 Log Management

Die in informationstechnischen Systemen während der Erstellung von neuen Prozessen entstehenden Daten oder Änderungen an diesen Daten, die durch bestehende Prozesse bedingt sind, werden nur selten vollständig protokolliert. Im Bereich der Computer- und Netzwerksicherheit ist dieses Vorgehen jedoch notwendig, um Anomalien im Betrieb sowie Angriffsmuster und deren Ursprung nachvollziehen zu können. Die hierzu relevanten Datenbestände, die im produktiven Betrieb eines informationstechnischen Systems entstehen, müssen daher zentral vorgehalten und ausgewertet werden. Bei der Erfassung und Verarbeitung der Daten gilt es, elementare Herausforderungen, die sich aus der Komplexität und der Quantität der Daten ergeben, zu lösen. Eine manuelle, textgestützte Analyse dieser Daten auf Systemebene, um Anomalien, Fehlkonfigurationen und aktive bzw. versuchte Angriffe zu identifizieren, ist bei hochfrequentierten Ereignissen nicht möglich. Daher müssen sicherheitsrelevante Infrastrukturkomponenten identifiziert und an ein zentrales Log-Management angebunden werden (Chuvakin et al. 2013, S. 29-32).

Um Kontextinformationen zu den Auslösern möglicher Anomalien zu erhalten, ist es mitunter auch notwendig, Logdaten von Netzwerkteilnehmern zu protokollieren. Somit sind nicht nur sicherheitsrelevante Komponenten der Kerninfrastruktur von der Ereignisprotokollierung betroffen, sondern auch Endgeräte, wie Desktops, Laptops, andere tragbare Geräte und Anlagen wie netzwerkfähige Drucker und Telefone. Die

Datenbestände können dann zentral gespeichert und überwacht werden. Die Informationsvisualisierung kann dabei helfen, die Daten inhaltlich und strukturell aufzubereiten, sodass Anomalien, Fehlkonfigurationen und aktive Angriffe identifiziert werden können. Abbildung 9 enthält eine Sammlung von Visualisierungen, ein Dashboard, in dem mehrere Logquellen dargestellt sind.

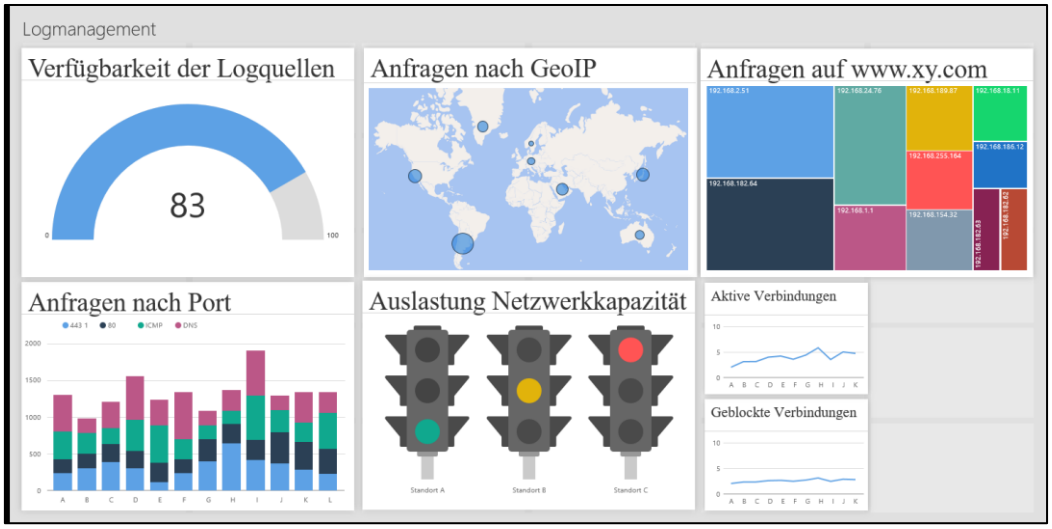

Abbildung 9: Logmanagement Dashboard (Eigene Darstellung)

Bei der Analyse von Auffälligkeiten mittels Dashboard handelt es sich um eine explorative Art der Visualisierung, die es dem Benutzer erlaubt, die dargestellten Inhalte anzupassen. Die Identifizierung von Anomalien und Fehlerzuständen in den betroffenen Systemen oder Netzwerken ist somit ein manueller Prozess, in dessen Verlauf die Erkennungsrate von den Fähigkeiten und dem Wissen des Benutzers abhängt. Daher sollten die verfügbaren Datenbestände aus den Quellsystemen in ein Security-Information-and-Event-Management-System (SIEM-System) überführt werden. Dieses ermöglicht die Definition von Regelsätzen, mit denen wiederum Anomalien und Fehlerzustände automatisiert identifiziert werden und einen zu definierenden Adressatenkreis mittels einer, den Vorfall beschreibenden, Ereignismeldung informieren (Miller 2011, S. 77-93).

Visualisierte Informationen werden jedoch auch in diesen Systemen dazu verwendet, Fehlerzustände initial zu erkennen und auf Basis der dadurch gewonnenen Erkenntnisse die entsprechenden Regelsätze zur Erkennung dieser Vorfälle zu definieren. Auf dieser Basis können Ereignisse zeitnah mit anderen relevanten Ereignissen verknüpft werden.

Informationen über Beziehungen können dann erneut durch eine Visualisierung beschrieben werden, sodass es möglich ist, die Ursache eines Vorfalls zu identifizieren und reaktive Maßnahmen zur Behebung dieses Vorfalls einzuleiten. Beispiele für Anomalien in Betriebs- und Anwendungssystemen oder Netzwerkkomponenten variieren dabei je nach Hersteller und Einsatzzweck. Generische Fehlerzustände und Anomalien sind z. B. Vorfälle, bei denen privilegierte Berechtigungen ohne Autorisierung genutzt werden oder der Versuch registriert wird, eine entsprechende Authentifizierung zu umgehen. Analog hierzu lassen sich unberechtigte Verbindungsversuche, die Verwendung von nicht unterstützten Protokollen oder Daten- und Informationsabflüsse über vorhandene Schnittstellen nennen. Die Übersichtsfunktion der Visualisierung beschränkt die Analysemöglichkeiten dabei zunächst auf offensichtliche Anomalien und Fehlerzustände, die trotz ihrer Prägnanz in einer textuellen Analyse nur schwer ausfindig zu machen wären. Erst durch die Funktion, Inhalte bewusst zu selektieren oder auszuschließen und damit den Suchraum einzugrenzen, ist es möglich, auch unauffällige Abweichungen oder komplexe Ereignisse zu identifizieren. Die Visualisierung von Ereignisprotokollen kann demnach dazu genutzt werden, Maßnahmen zu identifizieren, mit denen die Funktion und Betriebssicherheit von Anwendungs- und Betriebssystemen und Netzwerkkomponenten sichergestellt werden (Krasser et al. 2005, S. 42-49). Gleichzeitig können auf Basis von Informationen, die aus der Visualisierung abgeleitet werden, präventive Maßnahmen zum Schutz der eingesetzten Komponenten und der Benutzer implementiert werden. Sofern ein Vorfall aufgetreten ist, der als zivil- oder strafrechtlicher Verstoß gilt, ist es möglich, die Ursache und den Tatverlauf visuell aufzuarbeiten und die Ergebnisse an eine dafür zuständige Ermittlungsbehörde weiterzuleiten.

4.3.2 eDiscovery

Electronic Discovery (eDiscovery) ist ein Verfahren, das im Rahmen von zivil- oder strafrechtlichen Ermittlungen eingesetzt wird, um elektronische Beweise zu erfassen und aufzubereiten. Dabei werden Daten und elektronisch gespeicherte Informationen, die sich aus diesen Daten ergeben, zuerst identifiziert, um sie im Anschluss in einem mehrstufigen Prozess zu verarbeiten. Für das Vorgehen vor, während und nach der Verarbeitung dieser elektronisch gespeicherten Informationen (Electronically Stored Information) existiert ein Referenzmodell mit dem Titel ‚Electronic-Discovery-Reference-Model' (EDRM). In

Abbildung 10 werden die Phasen dieses eDiscovery-Referenzmodells dargestellt (EDRM Duke Law 2019).

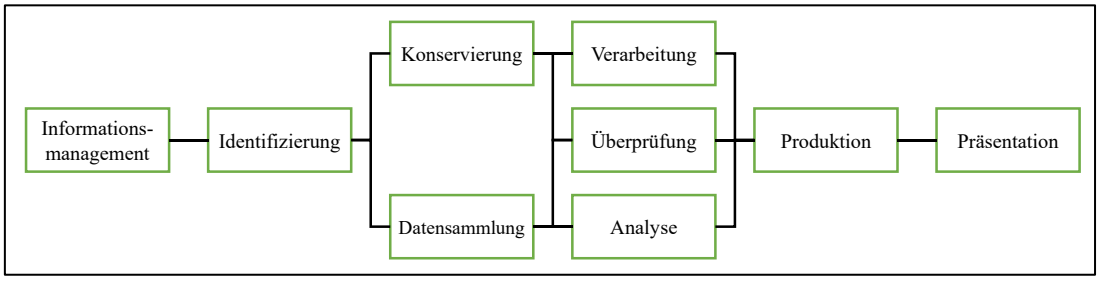

Abbildung 10: Phasen des Electronic-Discovery-Reference-Model (Eigene Darstellung, in Anlehnung an: (EDRM Duke Law 2019))

Das EDRM beschreibt eine konzeptionelle Sicht auf den eDiscovery-Prozess. Das heißt, dass nicht alle der beschriebenen Teilprozesse ausgeführt werden müssen und führt dazu, dass der Prozess im Bedarfsfall angepasst werden kann. Das Referenzmodell sieht die folgenden Aufgaben und Voraussetzungen für die einzelnen Prozessphasen vor (EDRM Duke Law 2019):

- Informationsmanagement:
  Diese Phase umfasst die Verwaltung von Risiken und des Ablageorts einer Datei. Die Verwaltung der Datei durch das Informationsmanagement erfolgt über die gesamte Lebensdauer der Datei, ab dem Zeitpunkt der Erstellung bis zur endgültigen Löschung.
- Identifizierung:
  In dieser Phase wird bestimmt, aus welcher Quelle die elektronisch gespeicherte Information gesichert werden soll.
- Konservierung:
  Hier werden die Maßnahmen definiert, die ergriffen werden müssen, um die Datei vor unberechtigten Veränderungen oder Zerstörung zu schützen.
- Datensammlung:
  Bei der Datensammlung geht es um die Sicherung der elektronisch gespeicherten Informationen zur weiteren Verwendung im eDiscovery-Prozess.
- Verarbeitung:
  Im Rahmen der Verarbeitung werden Maßnahmen zur Normierung und Optimierung für die weiteren Prozessphasen bestimmt.

- Überprüfung:
  Die elektronisch gespeicherten Informationen werden auf ihre Relevanz im Rahmen der Vorfallsbearbeitung geprüft.
- Analyse:
  Bei der Analyse werden Maßnahmen zur Bewertung der elektronisch gespeicherten Informationen anhand von Faktoren, mit denen Inhalt und Kontext der Information beschrieben werden können, ergriffen.
- Produktion:
  Die elektronisch gespeicherten Informationen werden für Prozessbeteiligte bereitgestellt.
- Präsentation:
  Zuletzt werden die elektronisch gespeicherten Informationen vor möglichen Zielgruppen präsentiert. Der Fokus liegt auf einer adressatengerechten Aufbereitung, um Erkenntnisse zu beschreiben und bestehende Fakten zu validieren.

Mit der Eigenschaft des Prozesses, sämtliche elektronisch gespeicherten Informationen unabhängig von Form, Inhalt und Struktur untersuchen zu können, gehen auch Herausforderungen hinsichtlich der Identifikation und Bewertung der Informationen einher. Die visuelle Darstellung der vorhandenen Informationen bietet grundsätzlich die Möglichkeit, große Datenmengen und deren Inhalte sicht- und vergleichbar zu machen (Nordbø 2014, S. 86-92). Daher werden Methoden und Werkzeuge zur Informationsvisualisierung in den Prozessphasen der Identifizierung, Verarbeitung, Analyse und Präsentation verwendet, um Inhalte adressatengerecht aufzubereiten. Die folgenden Abbildungen 11 und 12 zeigen die Verwendung einer Visualisierung im Kontext einer anlassbezogenen Untersuchung zur Sammlung von Personal-Store-Dateien (PST).

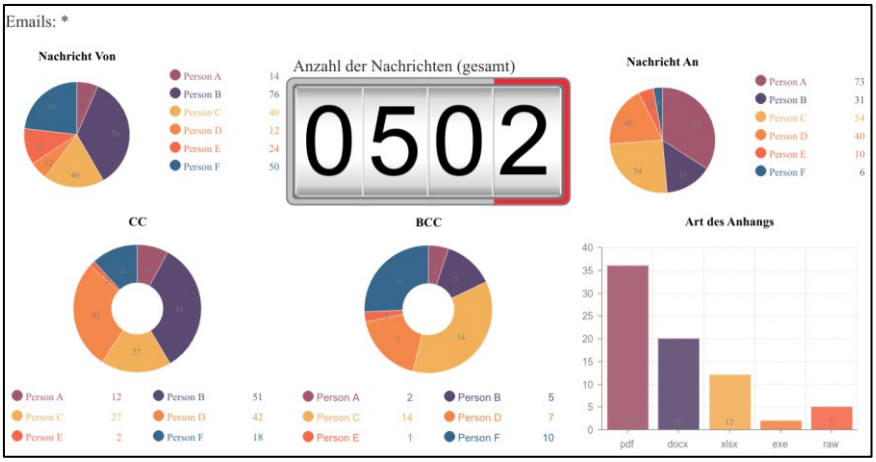

Abbildung 11: eDiscovery – alle Elemente einer Personal-Store-Datei (Eigene Darstellung)

Abbildung 11 umfasst eine Übersicht über alle verfügbaren E-Mail-Nachrichten, die im Rahmen der Datensammlung sichergestellt werden konnten. Im nächsten Prozessschritt, der Verarbeitung, können diese Daten aufbereitet und auf Vollständigkeit überprüft werden. Hierbei kann es zur erneuten Ausführung der Datensicherung kommen, sofern die während der Datensicherung gespeicherten Inhalte von denen der zuvor, in der Phase der Identifikation, ermittelten Datenbeständen abweichen. Abbildung 12 beschreibt das Resultat einer Abfrage während der Analysephase. Hierbei wurden Nachrichten von den Personen A und D sowie die Dateinamenserweiterungen ‚pdf' und ‚exe' ausgewählt.

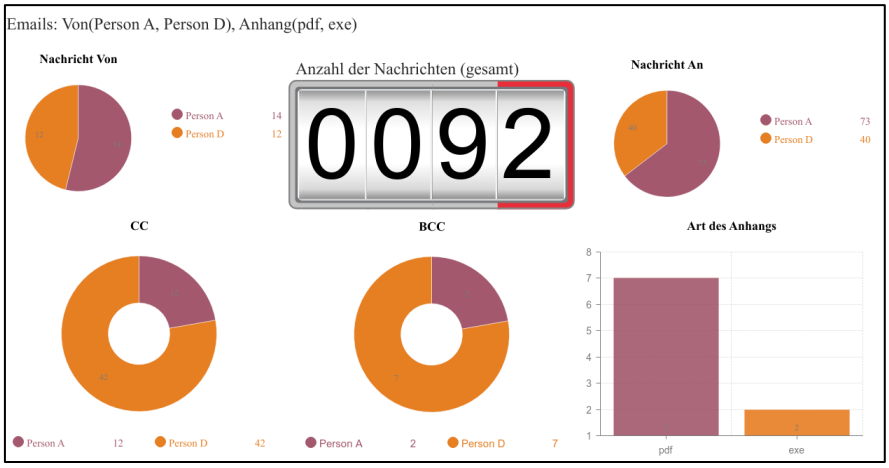

Abbildung 12: : eDiscovery – ausgewählte Elemente einer Personal-Store-Datei (Eigene Darstellung)

Anhand dieses Beispiels kann gezeigt werden, dass die Informationsvisualisierung während der Durchführung verschiedener Phasen des eDiscovery-Prozesses genutzt werden kann, um Inhalte verständlich darzustellen und auf dieser Basis tiefgreifendere

Ansatzpunkte, hinsichtlich der Ursache oder zur Aufklärung eines Vorfalls, zu identifizieren. Die Möglichkeiten bzgl. der Gestaltung der Visualisierung sind grundsätzlich nur durch die verfügbare Datenbasis beschränkt.

So ist es auf Basis von PST-Dateien möglich, eine historisierte Beschreibung des Nachrichtenverlaufs zu erstellen und mögliche Ursachen für Veränderungen dieses Verlaufs zu ermitteln. Hierzu können Ereignisse in Relation zu anderen Ereignissen gesetzt werden, um Muster und Abhängigkeiten zwischen den Kommunikationspartnern zu identifizieren. Die Visualisierung bietet damit eine effektive Möglichkeit, Betrachter schnell in Themenfelder und Vorfälle einzuführen, über deren Entstehung und Verlauf sie zuvor nicht informiert waren. Hierzu kann eine Reihe verschiedener Visualisierungen entworfen werden und diese können gegeneinander getestet werden, um zu bestimmen, welche der Visualisierungen das Szenario inhaltlich am wirksamsten beschreibt.

5 Visualisierungsmethoden und Werkzeuge zur forensischen Datenanalyse

Die von Strafverfolgungsbehörden und Unternehmen eingesetzten Werkzeuge und Methoden zur forensischen Datenanalyse sind häufig in Anwendungspaketen gebündelt. Diese Anwendungen werden vor allem dann eingesetzt, wenn ein prozessbeteiligter Ermittler oder Analyst über eine Kopie der auf einem Endgerät gespeicherten Daten verfügt, eine Untersuchung von Protokoll- oder Anwendungsdateien vornehmen soll oder die Ursächlichkeit ausgewählter Ereignisse für einen Vorfall bewerten muss. Diese Daten werden durch die Anwendung erfasst, verarbeitet und kategorisiert, sodass sie durch den Benutzer der Anwendung analysiert werden können. Das Format der Darstellung variiert dabei in Abhängigkeit von der gewählten Datenkategorie. Die Werkzeuge, unabhängig davon, ob sie kommerzielle oder Open-Source-Anwendungen sind, bieten Ermittlern und Analysten die Möglichkeit, die erfassten Datenbestände inhaltlich vollständig zu analysieren.

In Bezug auf die praktischen Einsatzmöglichkeiten erfüllen diese Anwendungen zentrale Anforderungen an die forensische Datenanalyse (Osborne et al. 2012, S. 35-47). Funktionen, wie die vollständige Protokollierung der ausgeführten Analyseschritte oder die Teilautomation des Analyseprozesses durch eine Vorverarbeitung der Datenbestände, bieten weitreichende Möglichkeiten während der Analyse und der Dokumentation. Das gilt vor allem bei der Identifikation von wiederkehrenden und neuen Tatbeständen in informationstechnischen Systemen. Das zur Identifikation von Tatbeständen und Anomalien eingesetzte Vorgehen basiert auf aussagekräftigen Mustern und Korrelationen innerhalb der bestehenden Daten, die es im Rahmen des Analyseprozesses zu erkennen gilt. Auf Grundlage der Erkenntnisse des Analyseprozesses ist es möglich, präventive Maßnahmen zur Vermeidung von unrechtmäßigen oder kriminellen Täuschungshandlungen zu implementieren, sodass die damit verbundenen Risiken reduziert werden können.

Vor diesem Hintergrund erfolgt eine forensische Datenanalyse häufig anlassbezogen und ist stets Teil einer technischen Fehleranalyse, einer Untersuchung von zivil- und strafrechtlichen Verstößen, einer Prozessüberprüfung, Revisionstätigkeiten oder anderen Sonderprüfungen. Die Form und damit auch die Verarbeitungsmöglichkeiten der

Rohdaten bestimmen die verfügbaren Analysemethoden und -werkzeuge. Der Strukturierungsgrad der Rohdaten kann dazu führen, dass die im Rahmen einer weiterführenden Analysetätigkeit benötigten Datenbestände transformiert, d. h. strukturell und inhaltlich aufbereitet werden müssen. Die Transformationsprozesse werden dabei vollständig dokumentiert, sodass sie bei Bedarf reproduziert werden können. Dieses Vorgehen lässt sich auf alle Formen von elektronisch gespeicherten Informationen anwenden, um bspw. Protokoll- und Kommunikationsdaten für die Datenanalyse aufzubereiten. Aus den anlassbezogenen Untersuchungen lassen sich Methoden und Werkzeuge zur forensischen Datenanalyse sowie Einsatzgebiete ableiten, die in den folgenden Unterkapiteln vorgestellt werden sollen.

5.1 Informationsvisualiualisierung als Hilfsmittel während der forensischen Datenanalyse

Maßnahmen und Hilfsmittel, die dazu beitragen, dass Informationen für prozessbeteiligte Personen verständlich aufbereitet werden können, gewinnen zunehmend an Bedeutung. Die Datenvisualisierung wird deshalb verwendet, um Ereignisse und Vorfälle zu beschreiben oder zu rekonstruieren. Dabei wird eine Information visualisiert, deren zugrunde liegenden Daten digital aufgezeichnet oder digitalisiert wurden. Die Beschreibung zeigt auf, was wann wie und wo passiert ist und wer es getan hat. Zudem ermöglicht sie einen Ausblick auf die Antwort zu der Frage, was getan werden kann, um sicherzustellen, dass ein Ereignis zukünftig eintritt oder ausbleibt (Baryamureeba und Tushabe 2004, S. 1-9). Aus diesem Grund ist die Visualisierung von Datenbeständen während der forensischen Datenanalyse eine zentrale Komponente, die es ermöglicht, Abweichungen und Muster zu erkennen und Daten anlassbezogen zu analysieren (Glatz 2010, S. 58-63; Fischer 2016, S. 9-36). Mit den zur Informationsvisualisierung eingesetzten Methoden und Werkzeugen werden große Datenmengen für einen Analysten oder Ermittler in einer Form aufbereitet, die es ermöglicht, Anpassungen an der gewählten Darstellung vorzunehmen (Tominski und Aigner 2015). Diese Option besteht unabhängig vom Quellsystem der Daten und führt dazu, dass der Analyst bzw. Ermittler einen Überblick über die Daten gewinnt und diesen Einblick im Kontext der Untersuchung einbringen kann, um relevante Inhalte zielgerichtet zu identifizieren.

5.2 Verwendungskontext der Methoden und Werkzeuge

Um Datenbestände und Informationen im Rahmen der forensischen Datenanalyse bewerten zu können, existieren verschiedene Werkzeuge und Methoden. Der Einsatz dieser Hilfsmittel erfolgt unter der Prämisse, dass deren Verwendung einen positiven Einfluss auf die Verständlichkeit der Daten hat und die Analysetätigkeit infolgedessen effektiver durchgeführt werden kann (Phan 2007, S. 79-82). Vor diesem Hintergrund sind alle Maßnahmen, die dazu führen, dass Daten und Informationen für den Analysten oder andere prozessbeteiligte Personen verständlicher dargestellt werden können, ein Hilfsmittel während der Untersuchung eines Vorfalls.

Ein Beispiel hierfür ist die Darstellung von Dateisysteminformationen in Form einer Treemap-Struktur: In einer Untersuchung wurde gezeigt, dass die Benutzer dieser Visualisierungsform gesuchte Daten innerhalb des Dateisystems im Durchschnitt 35 Prozent schneller finden konnten als mit klassischen Befehlszeilentools (Teelink und Erbacher 2006, S. 71-75). Das bedeutet nicht, dass Befehlszeilentools während der forensischen Datenanalyse nicht als Hilfsmittel verwendet werden können. Es zeigt aber auf, dass im Kontext der Analysetätigkeit ggf. andere Methoden und Hilfsmittel zur Verfügung stehen, die eine wirksamere Analyse der gleichen Datenbestände ermöglichen (Staheli et al. 2014). Daher gilt es, Werkzeuge und Methoden zu identifizieren, die im Hinblick auf die zu analysierende Datenbasis den größtmöglichen Nutzen für die prozessbeteiligten Personen bieten. Das Untersuchungsergebnis gibt einen Ausblick auf den allgemeinen Nutzen der Informationsvisualisierung bei der Unterstützung kognitiver Prozesse und unterstreicht damit den potenziellen Wert dieses Hilfsmittels während forensischer Untersuchungen.

In den folgenden Unterkapiteln werden daher mögliche Datenquellen und deren Struktur in Kombination mit verfügbaren Methoden und Werkzeugen beschrieben, die den Prozessbeteiligten im Kontext der Analyse eine effektive Arbeitsweise ermöglichen sollen. Der Fokus dieser Aufstellung liegt auf der für den Benutzer entstehenden Unterstützungsleistung durch das jeweils beschriebene Hilfsmittel in Bezug auf die Visualisierung von Daten und Informationen.

5.2.1 Ereignisprotokolle

Die Visualisierung von vorfallsrelevanten Ereignisprotokollen ist möglich, da diese im Regelfall aus Informationen wie einem Zeitstempel, einem oder mehreren Bezeichnern und Attributen, die das Ereignis beschreiben, bestehen. Aus diesen Informationen lassen sich in einer Ex-post-Betrachtung Ereignisse bestimmen, die einen Vorfall zur Folge hatten. Die Analyse von nutzergenerierten Daten, die während der Verwendung eines Systems oder einer Anwendung durch den Benutzer erzeugt werden, ist ein Anwendungsfall, bei dem es im Verlauf der Vorfallsbearbeitung zu hohen manuellen Aufwänden bei der Datenanalyse kommen kann. Ein- und ausgehende Netzwerkverbindungen oder Daten eines zentralen Proxy-Servers können während der forensischen Analyse eines sicherheitsrelevanten Vorfalls Informationen über den Verlauf und den Ursprung des Vorfalls liefern. Wurde ein Vorfall mittels dieser Datenquellen identifiziert, kann es notwendig sein, aufgezeichneten Hypertext-Transfer-Protocol-(HTTP-) und Hypertext-Transfer-Protocol-Secure-(HTTPS-) Netzwerkverkehr, der vor, bei und nach dem Vorfall stattgefunden hat, zu analysieren. Die Struktur dieser Daten kann durch Verweise und Weiterleitungen verschachtelt sein, wodurch eine textuelle Analyse zur Identifikation des schadhaften Netzwerkverkehrs komplex und zeitaufwendig ist. Um die Komplexität sowie den Aufwand während der Analyse solcher Vorfälle zu reduzieren, bietet die Anwendung Hviz die Möglichkeit, Netzwerkverkehr, der über die Protokolle HTTP und HTTPS übertragen wurde, auf einer Ereigniszeitachse zu visualisieren. In Abbildung 13 ist die Benutzeroberfläche der Anwendung Hviz dargestellt (Gugelmann et al. 2015).

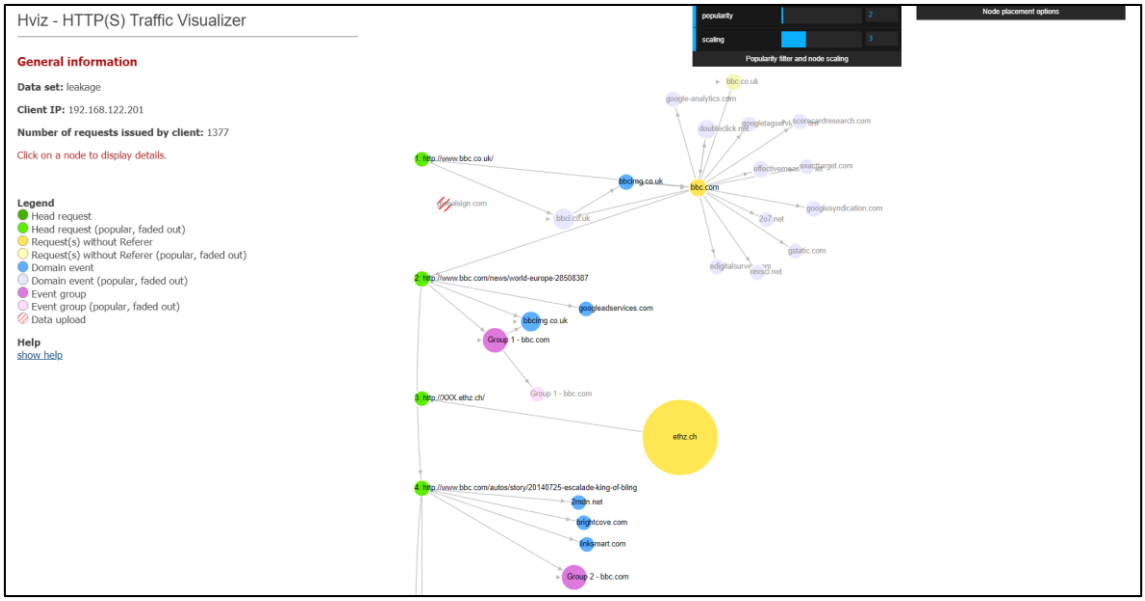

Abbildung 13: Visualisierung von HTTP- und HTTPS-Aktivitäten mit Hviz (Gugelmann et al. 2015)

Die historisierte Darstellung des Netzwerkverkehrs, der zum Zeitpunkt eines Vorfalls von einem Arbeitsplatz übertragen wurde, führt zu einer wahrnehmbaren Effektivitätssteigerung in der Analysephase. Durch die aggregierte und zeitlich strukturierte Darstellung wird der Ereignisverlauf verständlich aufbereitet. Die einzelnen Netzknoten enthalten zudem Kontextinformationen zu bestehenden Relationen. Auf dieser Grundlage ist es möglich, Vorfallsursachen, wie Netzwerkaktivitäten eines Schadcodes, Fehlkonfigurationen oder die Handlung eines Innentäters, zu identifizieren oder auszuschließen. Unter Berücksichtigung der durch die Analyse gewonnenen Erkenntnisse können reaktive Maßnahmen zur Eingrenzung des möglichen Schadens für andere Netzwerkteilnehmer eingeleitet werden. Hierzu kann das Ist-Schadenspotenzial einer Zieladresse überprüft werden. Eine Übersicht über die Zieladresse samt nachgelagerter Hintergrund-Aufrufe und implementierter Weiterleitungen kann über die Anwendung Lookyloo erstellt werden. In Abbildung 14 wird diese Übersicht für die Internetseite ‚www.bbc.co.uk' dargestellt (CIRCL.lu 2019).

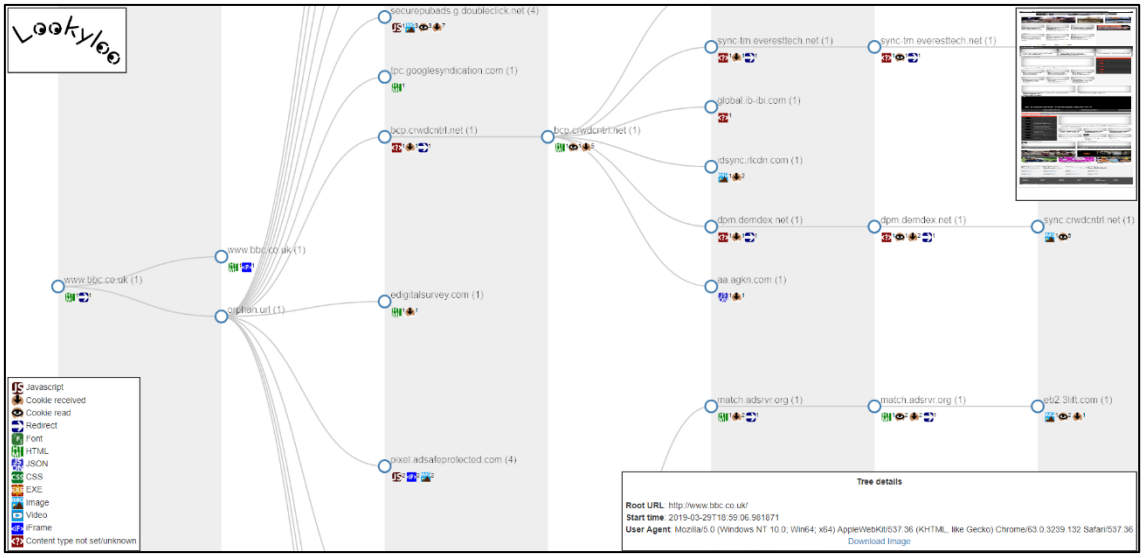

Abbildung 14: Überprüfung von ‚www.bbc.co.uk' mit Lookyloo (Eigene Darstellung, in Anlehnung an: (CIRCL.lu 2019))

Zu beachten ist hierbei, dass auf Basis der Informationen, die durch die Anwendung zur Verfügung gestellt werden, ausschließlich eine Ist-Bewertung des Schadenspotenzials der Zieladresse abgegeben werden kann. Eine historische Ansicht, bspw. für den Vorfallszeitpunkt, kann nicht erstellt werden.

5.2.2 Verbindungs- und Geodaten

Bei der Darstellung von vorfallsrelevanten Daten, die durch Verbindungen zwischen einem möglichen Täter oder Quellsystem und einem Betroffenen oder Zielsystem anfallen, kann es für die Verständlichkeit während der Analyse förderlich sein, wenn diese Informationen durch einen räumlichen und zeitlichen Zusammenhang ergänzt werden (Hargreaves und Patterson 2012, S. 69-79). Anwendungen, die dazu verwendet werden, Anrufprotokolle oder Einträge aus sozialen Medien und Standortdaten von mobilen Endgeräten zu korrelieren, bieten die Möglichkeit, Vorfallsinformationen in die entsprechende Darstellung zu integrieren. Die Anwendung GeoTime, die ausschließlich im Rahmen der Strafverfolgung eingesetzt werden kann, ermöglicht die korrelierte Darstellung mehrerer Datenquellen unter Zuhilfenahme der räumlichen und zeitlichen Beschreibung eins Ereignisses. Abbildung 15 zeigt eine mehrdimensionale Darstellung eines Vorfalls unter Berücksichtigung des Eintrittszeitpunkts und des Orts (GeoTime 2019).

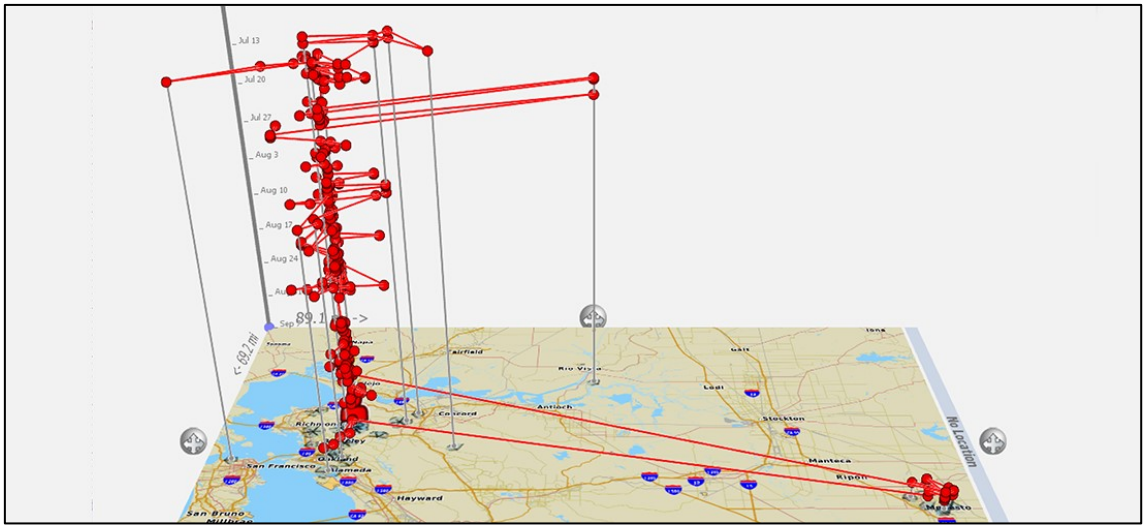

Abbildung 15: Ereignisvisualisierung unter Berücksichtigung des Eintrittszeitpunkts und Orts mit GeoTime (GeoTime 2019)

Die historisierte Darstellung von Verbindungs- und Geodaten kann in mehreren Phasen des forensischen Prozesses eingesetzt werden. Neben den Einsatzmöglichkeiten während der Analysephase, um Vorfälle und deren Ursache zu identifizieren, kann die Visualisierung auch in der Dokumentationsphase genutzt werden, um das Verhalten eines Beschuldigten oder den Systemzustand nach einem Vorfall darzustellen. Prozessbeteiligte Personen können hierdurch einen vollständigen Überblick über den Zeitverlauf eines Vorfalls gewinnen und bei Bedarf Ereignisse sowohl zeitlich als auch räumlich eingrenzen. Über die gekennzeichneten Schaltflächen innerhalb der Visualisierung von GeoTime wird eine explorative Navigation zum ausgewählten Zeitraum oder räumlichen Ausschnitt ermöglicht.

5.2.3 Dateisysteminformationen

Die Untersuchung auf das Vorhandensein einer Datei oder eines Dateiformats innerhalb eines Dateisystems kann textuell über einen Kommandozeilenbefehl realisiert werden. Komplexere Fragestellungen, wie die Überprüfung eines Ablageorts oder die Häufigkeit, mit der eine Datei innerhalb eines Dateiservers vorhanden ist, können über eine Visualisierung beantwortet werden. Durch das Verschlagworten (Indexierung) der Attribute des Dateisystems können die darin enthaltenen Informationen für die grafische Darstellung der Dateien genutzt werden. Abbildung 16 zeigt die Untersuchung eines Dateipfades innerhalb eines lokalen Dateisystems.

			Created 2018						
			Februar	März			April	Mai	Juni
Full path	Name		4	27	28	30	31	4	4	4
LoneWolf.E01 - Partition 4 (Microsoft NTFS, 476.34 GB)\Users..	AIRPORT INFORMATION.docx	Anzahl der Datensätze				1				
		Size (bytes)				172.684				
LoneWolf.E01 - Partition 4 (Microsoft NTFS, 476.34 GB)\Users..	AMEN.pdf	Anzahl der Datensätze								1
		Size (bytes)								371.701
LoneWolf.E01 - Partition 4 (Microsoft NTFS, 476.34 GB)\Users..	BladeofGrass.jpg	Anzahl der Datensätze					1			
		Size (bytes)					201.810			
LoneWolf.E01 - Partition 4 (Microsoft NTFS, 476.34 GB)\Users..	Box Sync.lnk	Anzahl der Datensätze			1					
		Size (bytes)			1.606					

Abbildung 16: Tabellarische Darstellung zur Analyse des Dateipfades (Eigene Darstellung)

Durch die Visualisierung werden Dateien dargestellt, die durch den im Testfall beschriebenen Nutzer vor dem Vorfall erstellt wurden. Auf Grundlage der angezeigten Attribute ist es möglich, die zu analysierende Datenmenge einzugrenzen und vorfallsrelevante Inhalte schneller zu identifizieren. Die Nutzung eines Filters in Kombination mit der Anzeige weiterer Attribute, in diesem Beispiel der MD5-Hash und der SHA1-Hash, wird in Abbildung 17 dargestellt.

				Accessed 2018		
				Februar	April	Mai
File attribu..	Name	MD5 hash	SHA1 hash	4	4	4
Archive	AIRPORT INFORMATION.docx	297eec248647f33f887d7..	e6e04a4d0c8c69a9fc40c..		■	
	Cloudy thoughts (4apr).docx	f8c2bc733c109a88405df..	308c3ca689f0e6603acb7..			■
	Planning.docx	4ef414e469b7830faa2db..	b03dd6ecd3272cb67476b..		■	
	The Cloudy Manifesto.docx	14c07920ddc81fbd489e6..	18f1486346325876 7d7d..	■		

Abbildung 17: Tabellarische Darstellung zur Analyse des Dateipfades, erweitert durch die Attribute MD5 und SHA1 (Eigene Darstellung)

Die durch die Visualisierung bestimmten Dateien können dann erneut inhaltlich verschlagwortet werden, um die verwendete Sprache oder den Inhalt über Stich- und Suchwortlisten zu analysieren. Eine ähnliche Form der Visualisierung kann dazu verwendet werden, um den Urheber von inkriminierten Inhalten auf einem zentralen Dateiserver zu bestimmen. Abbildung 18 beschreibt hierzu den Zeitverlauf, über den die Dateien von den einzelnen Benutzern im Dateisystem abgelegt wurden.

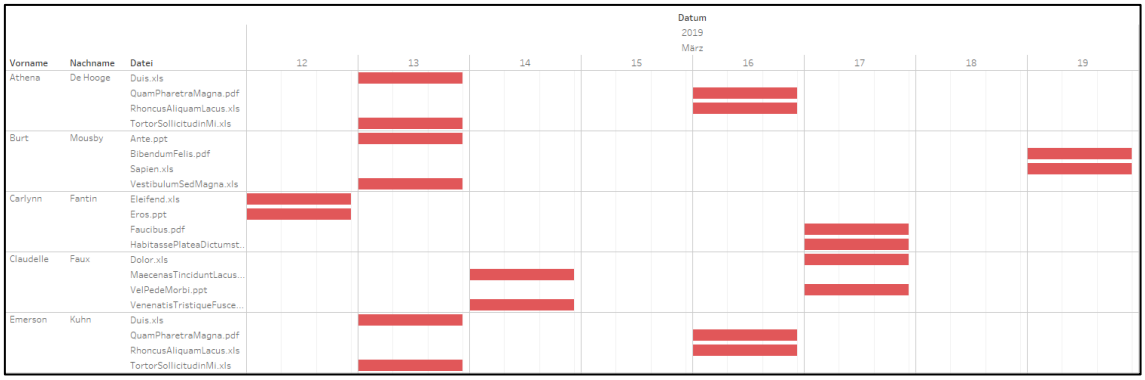

Abbildung 18: Zeitachse zur Ablage von inkriminierten Inhalten auf einem Dateiserver (Eigene Darstellung)

Auf diese Weise ist es möglich, alle Benutzer zu identifizieren, die eine der betroffenen Dateien in dem Dateisystem abgelegt haben. Die Daten- und Informationsvisualisierung bietet in diesen Fällen eine Möglichkeit, die bestehenden Datenmengen zielgerichtet und wirksam zu analysieren.

5.2.4 Textdaten

Die inhaltliche Analyse von Textdateien ist ein Bestandteil der forensischen Untersuchung. Textdateien werden hierbei als digitale oder digitalisierte Texte definiert. Diese sind mehrheitlich unstrukturiert, sodass die Analysetätigkeit ein zeitintensiver und zugleich weitgehend manueller Prozess ist (Al-Zaidy et al. 2011, S. 72). Die visuelle Darstellung einzelner vorfallsrelevanter Wörter oder Wortgruppen ermöglicht es dem Analysten oder Ermittler, eine Beziehung zwischen verschiedenen Informationen zu erkennen und diese zu beschreiben (Jankun-Kelly et al. 2009). Um Füllinformationen auszuschließen und verwendete Soziolekte zu identifizieren, ist es möglich, eine Datenhierarchie einzuführen. Eine solche Hierarchie wird in Form eines Dendrogramms für den Wortstamm ‚ford' und ‚forsch' in Abbildung 19 dargestellt.

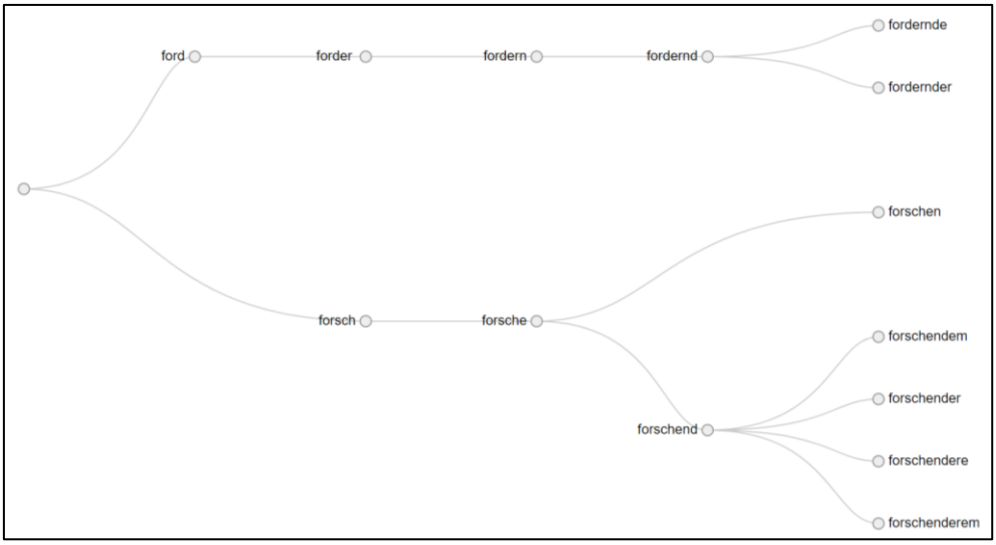

Abbildung 19: Datenhierarchie zur Analyse von Wortgruppen (Eigene Darstellung)

Diese Darstellung ermöglicht es, die Verwendung, die Struktur und mögliche Anomalien im Aufbau einzelner Wortgruppen zu identifizieren. Die Grundlage dafür bildet ein frei wählbarer Wortstamm, in Abbildung 19 ‚ford' bzw. ‚forsch'. So kann die Wahrscheinlichkeit reduziert werden, dass vorfallsrelevante Informationen in Textdateien, z. B. durch die Verwendung eines Jargons oder Soziolekts unberücksichtigt bleiben. Auf Grundlage dieser Information können Such- und Stichwortlisten erstellt werden. Einzelne Wortbestandteile innerhalb dieser Liste können durch reguläre Ausdrücke oder definierte Variablen ergänzt werden, sodass auch syntaktische Schwächen innerhalb des Textes berücksichtigt werden (Lang Beebe et al. 2011, S. 732-744; Schwartz und Liebrock 2008). Der Einsatz von Wortwolken in Verbindung mit einer Freitextsuche ermöglicht es, die Satzposition einzelner Such- und Stichwörter zu bestimmen. Diese Maßnahmen zur Daten- und Informationsvisualisierung von Textdaten können Analysten und Ermittler während einer forensischen Untersuchung bei der Identifikation relevanter Dokumente und Textabschnitte unterstützen.

5.3 Funktion und Nutzen eingesetzter Methoden und Werkzeuge

Die während der forensischen Datenanalyse eingesetzten Visualisierungsmethoden und -werkzeuge können auch in anderen Phasen des forensischen Prozesses unterstützend eingesetzt werden. Die Eigenschaft, Daten und deren Abhängigkeiten auf Grundlage einer Übersicht zu bewerten und ausgehend davon einzelne oder Gruppen von Dateien

auszuwählen, um sie inhaltlich vollständig zu analysieren, ermöglicht eine zielgerichtete Datenanalyse. Der Nutzen der Visualisierung besteht dabei auch in der Dokumentation der Untersuchungsergebnisse. Das während der Analyse gewählte Vorgehen ist allgemein verständlich und ermöglicht es, die Erkenntnisse zielgruppengerecht aufzubereiten. Eine inhaltliche Analyse der Daten durch den Anwender rückt damit in den Vordergrund des forensischen Prozesses (Labudde und Spranger 2017, S. 7-20). Um technische Hürden, die bei der Erstellung und im Betrieb dieser Hilfsmittel aufkommen, zu reduzieren, können der Aufbau und der Inhalt möglicher Quelldaten erfasst werden. Für diese Strukturen können geeignete Darstellungsformen bestimmt werden, wodurch fallbasierte Aufwände zur Bestimmung der Visualisierungsmöglichkeiten reduziert werden. Die Zielsetzung vor und während der Auswahl von Methoden und Werkzeugen zur Datenvisualisierung sollte dabei stets die Bestimmung von nutzerfreundlichen Mensch-Technik-Schnittstellen sein, da diese die Grundlage für die effektive Nutzung von Visualisierungen während des forensischen Prozesses bilden. Durch das gewählte Vorgehen, bei dem Daten aus den Quellsystemen extrahiert werden, ist es möglich, die Analysetätigkeit unabhängig von Systemspezifikationen durchzuführen. Diese Abstraktion ermöglicht eine zielgerichtete Analyse von großen Datenbeständen in aggregierter Form. Die Informationsgewinnung auf Basis der erstellten Visualisierungen sollte intuitiv möglich sein, sodass die Resultate einem breiten Adressatenkreis zugänglich sind und sie auch in der Abschlussdokumentation eingesetzt werden können.

6 Anforderungen der forensischen Datenanalyse an die Daten- und Informationsvisualisierung

An die forensische Datenanalyse und ihre Umsetzung werden zentrale Anforderungen gestellt, die sich unter anderem aus dem IT-Grundschutz des BSI ergeben (BSI 2019). Diese gelten für alle Teilgebiete der IT-Forensik und damit auch für die während der Untersuchung und Datenanalyse eingesetzten Werkzeuge und Methoden. Im Verlauf der Analysephase werden vorfallsrelevante Daten sowohl isoliert als auch im Gesamtzusammenhang des Vorfalls betrachtet. Bei den dazu verwendeten Werkzeugen muss sichergestellt werden, dass deren Funktion gegeben ist und sie nicht während der Analysetätigkeit manipuliert werden können. Des Weiteren sollte es allen prozessbeteiligten Ermittlern und Analysten möglich sein, die Methode oder das Werkzeug anzuwenden. Aus diesen und weiteren Anforderungen ergeben sich Herausforderungen, die es bei der Verwendung von interaktiven Grafiken und Visualisierungen während der forensischen Datenanalyse zu beachten gilt. Daher werden in den folgenden Unterkapiteln vorbereitende und operative Anforderungen des forensischen Prozesses sowie deren Auswirkungen auf die visuelle Darstellung von Informationen während der forensischen Datenanalyse beschrieben.

6.1 Allgemeine Anforderungen an die IT-Forensik

An den forensischen Prozess und die während dieses Prozesses ausgeführten Phasen werden, unabhängig vom gewählten Vorgehensmodell, zentrale Anforderungen gestellt, mit denen sichergestellt wird, dass die Analyseergebnisse auch im Rahmen einer zivil- und strafrechtlichen Untersuchung genutzt werden können (Dewald et al. 2015, S. 63-114). Mit den folgenden sechs Punkten werden diese Anforderungen an die IT-Forensik beschrieben:

1. Akzeptanz:
   Die während der Untersuchung ausgeführten Maßnahmen und die dazu verwendeten Methoden und Werkzeuge müssen in der Fachwelt anerkannt sein und dokumentiert werden.

2. Glaubwürdigkeit:
   Die Funktion und die Robustheit der eingesetzten Methoden und Werkzeuge müssen sichergestellt werden können, sodass die durch sie ermittelten Analyseergebnisse glaubwürdig sind.
3. Wiederholbarkeit:
   Die während des forensischen Prozesses ermittelten Untersuchungsergebnisse müssen reproduziert werden können. Das bedeutet, dass, sofern die in der Dokumentation beschriebenen Untersuchungsschritte inhaltlich wiederholt werden, das Analyseergebnis identisch ist.
4. Integrität:
   Während der Analysetätigkeit werden die zu analysierenden Daten nicht verändert. Müssen Daten im Rahmen der Analyse transformiert werden, um sie bspw. inhaltlich bewerten zu können, ist es notwendig, dass ein Abbild der Originaldaten vorgehalten wird.
5. Ursache und Auswirkungen:
   Ziel des forensischen Prozesses muss es sein, Verbindungen zwischen verschiedenen Ereignissen herzustellen, da diese in Summe den Vorfall beschreiben. Wenn möglich, sollte bei der Bestimmung von bestehenden Verbindungen auch ein Ursprung, ggf. mit Personenbezug, hergestellt werden.
6. Dokumentation:
   Über die Phasen des forensischen Prozesses hinweg muss eine chronologische Dokumentation aller Erkenntnisse und damit verknüpft auch der ermittelten Beweise erstellt werden.

Der Einhaltung dieser Anforderungen sollte im Rahmen des forensischen Prozesses eine hohe Aufmerksamkeit zuteilwerden, da die Nichterfüllung ggf. zur Folge hat, dass die Untersuchungsergebnisse nicht zur zivil- und strafrechtlichen Bewertung eines Vorfalls verwendet werden können (BSI 2011, S. 22-33). Während und nach der Durchführung der einzelnen Phasen kann eine Erfolgsmessung in Bezug auf die allgemeinen Anforderungen an den forensischen Prozess erfolgen. Eine Möglichkeit hierfür besteht in der normierten Dokumentation ausgeführter Analyseschritte für bestimmte Vorfallsklassen: Diese ermöglicht zum einen eine reduzierte Einarbeitungszeit der prozessbeteiligten Ermittler und Analysten und stellt zum anderen sicher, dass die Anforderungen an die Integrität, Wiederholbarkeit und Glaubwürdigkeit der gewählten

Maßnahmen eingehalten werden können. Dieser lückenlose Nachweis über vorgenommene Untersuchungen an den Datenbeständen (engl. chain of custody) ermöglicht es auch, beginnend mit dem Zeitpunkt der Erfassung, einen Missbrauch oder Manipulationen an den Beweismitteln zu identifizieren (Hoelz und Ralha 2013).

Um diese Anforderungen an den forensischen Prozess und seine Untersuchungsergebnisse einhalten zu können, müssen diese abstrakten Vorgaben für die einzelnen Prozessphasen näher bestimmt werden. In den folgenden Unterkapiteln werden daher die organisatorischen und technischen Anforderungen an die Untersuchungs- und Analysephase konkretisiert.

6.2 Organisatorische Anforderungen

Die Umsetzung der Maßnahmen, die sich aus den organisatorischen Anforderungen ergeben, stellt in Verbindung mit den technischen Anforderungen sicher, dass alle ermittelten Erkenntnisse und Nachweise das Resultat einer forensisch korrekten Vorgehensweise während der Analyse der Datenbestände sind. Erfolgt die Datenanalyse in Teilen unter Zuhilfenahme von Methoden der Daten- und Informationsvisualisierung, ist es von Bedeutung, dass die Integrität der Datenbestände nicht verletzt und damit auch die Glaubwürdigkeit der Analyseergebnisse nicht beeinträchtigt wird (BSI 2011, S. 22-25). Hierzu muss das Verhalten der verwendeten Methoden und Werkzeuge zur Visualisierung im Grenzbereich, d. h. bei Unvollständigkeit der Datengrundlage und volumenbedingten Darstellungsgrenzen, bestimmt werden. Sofern es nicht möglich ist, das Verhalten der gewählten Hilfsmittel unter diesen Umständen zu bestimmen, sind die Glaubwürdigkeit und die Wiederholbarkeit der Analyseergebnisse gefährdet.

Für die Schnittmenge der organisatorischen und technischen Anforderungen an die verfügbaren Methoden und Werkzeuge muss bestimmt werden, welche Maßnahmen eingeleitet werden können, um deren Wirksamkeit während der Analysetätigkeit sicherzustellen. Diese Maßnahmen können eine Vorverarbeitung der Datenbasis beinhalten. Werden Datenbestände normiert oder transformiert, besteht der organisatorische Anspruch, dass die durchgeführten Maßnahmen und deren Einfluss auf die Datenbasis dokumentiert werden. Ist es nicht möglich, die zu analysierenden Daten durch eine Vorverarbeitung, hinsichtlich der Relevanz der Daten für den zu

analysierenden Vorfall, zu selektieren oder unvollständige Inhalte sachgerecht zu rekonstruieren, so müssen andere Vorgehensweisen für die Bewertung und Analyse der Daten in Betracht gezogen werden. Da die Datenanalyse einen Übergabepunkt bildet, bei dem Erkenntnisse aus der Vorfallsbearbeitung dazu verwendet werden, die zivil- oder strafrechtliche Verfolgung eines Vorfalls anzustoßen, müssen die ermittelten Nachweise so aufbereitet werden, dass sie nicht missinterpretiert werden können (Geschonneck et al. 2014, S. 98-100). Dies gilt insbesondere für grafisch dargestellte Informationen, die vorfallsrelevante Ereignisse und deren Ursache beschreiben. Kann ein Ereignis und dessen Ursache durch eine Visualisierung nicht selbsterklärend dargestellt werden, ist es möglich, die grafischen Darstellungen, die den Vorfall beschreiben, durch textuelle Beschreibungen bzgl. des Vorgehens und der daraus resultierenden Erkenntnisse aufzubereiten. Eine solche Informationsgrafik umfasst genaue Aussagen über die Analyseergebnisse und unterstützt prozessbeteiligte Personen bei der weiteren Vorfallsbearbeitung. Hierzu werden die Ereignisse entweder ausgehend von der ermittelten Ursache oder rückwirkend auf Basis des Vorfalls miteinander verknüpft. In Abbildung 20 wird eine Informationsgrafik dargestellt, die den Diebstahl und die Weitergabe geistigen Eigentums durch eine verdächtige Person beschreibt.

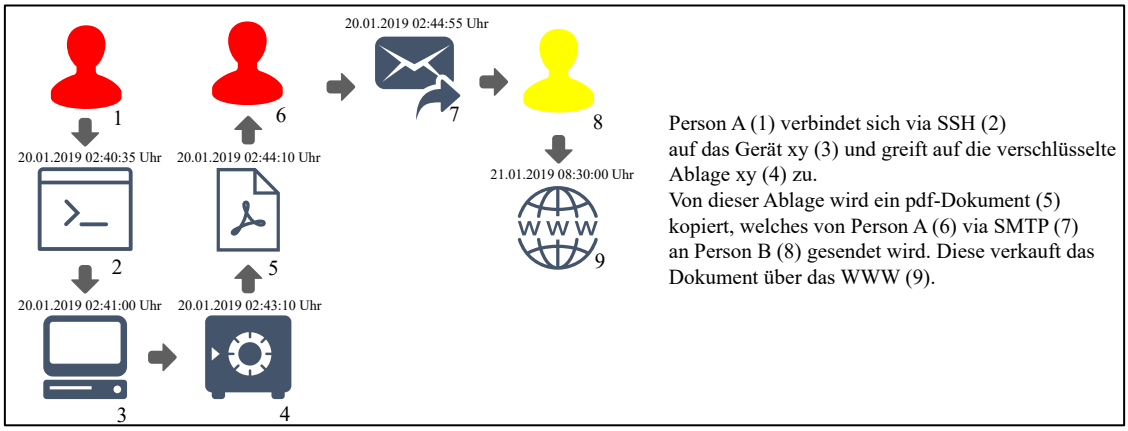

Abbildung 20: Informationsgrafik zur Vorfallsdarstellung (Eigene Darstellung)

Prozessbeteiligten Personen ist es auf dieser Grundlage möglich, einen Überblick über den gesamten Vorfall zu gewinnen. Durch die während der Untersuchung und Analyse erstellten Visualisierungen ist es zudem möglich, proaktiv oder auf Nachfrage grafisch aufbereitete Einzelnachweise zu den Analyseergebnissen zur Verfügung zu stellen. In Verbindung mit der Abschluss- und Prozessdokumentation werden so alle zuvor beschriebenen Anforderungen an den forensischen Prozess erfüllt.

6.3 Technische Anforderungen

Da die Resultate der forensischen Datenanalyse ein weiteres Vorgehen, auch hinsichtlich der rechtlichen Bewertung eines Vorfalls, bestimmen, ist es notwendig, dass die ermittelten Erkenntnisse aus einer vollständigen Analyse aller verfügbaren Datenbestände stammen. Daraus ergibt sich die Notwendigkeit, dass Datenbestände, die nicht aus den Quellsystemen extrahiert und daher auch nicht unter Verwendung von Methoden und Werkzeugen der Datenvisualisierung dargestellt werden können, durch andere Applikationen oder Methoden bewertet werden müssen. Bei Datenuntersuchungen und -analysen, die in Teilen unter Verwendung von interaktiven Grafiken oder anderen Formen der Daten- und Informationsvisualisierung durchgeführt wurden, sollte es daher möglich sein, die ermittelten Nachweise zu exportieren. Durch den Export der Analyseergebnisse in Verbindung mit Informationen zu den Maßnahmen, die ausgeführt wurden, um die Erkenntnisse zu gewinnen, ist es möglich, die Wirksamkeit der ausgewählten Methode oder des Werkzeugs aufzuzeigen (BSI 2011, S. 22-25). Durch die kontinuierliche Verwendung und Optimierung des Verfahrens ist es möglich, dass dieses Vorgehen in der Fachwelt anerkannt wird. Während der Durchführung der Analysetätigkeit muss sichergestellt werden, dass mögliche Abweichungen in den Systemzeiten verschiedener Quellsysteme normiert werden. Diese Verfahren sollten für alle Faktoren angewendet werden, deren Änderung keinen direkten Einfluss auf den Inhalt der zu analysierenden Daten hat. Die Normierung von verschiedenen Devisen, die während der Untersuchung eines Vorfalls aufkommen, auf eine zentrale Währungseinheit, wäre eine inhaltliche Anpassung der Daten und ist somit zu vermeiden. Anpassungen, die im Rahmen der Normalisierung vorgenommen werden, sollten auf den gesamten Datenbestand angewendet und vollständig dokumentiert werden. Dieses Vorgehen stellt sicher, dass eine nachträgliche Anpassung der Daten und damit auch eine Verletzung der Datenintegrität identifiziert werden können. Die Normierung bildet indes auch die Grundlage für die Identifikation von Vorfallsursachen, da sie es ermöglicht, Verbindungen zwischen Ereignissen aus verschiedenen Datenquellen herzustellen.

Um die Anforderungen an die Wiederholbarkeit und Glaubwürdigkeit der Untersuchung zu erfüllen, muss sichergestellt werden, dass ein lesender Zugriff auf die Datenbestände nur einem hierzu berechtigten Adressatenkreis möglich ist. Änderungen der Datenbestände sollten ausschließlich zu Transformationszwecken erfolgen. Ein Beispiel

hierfür sind syntaktische Anpassungen der Datenbestände (engl. parsing). Diese Maßnahmen sollten noch vor Beginn der eigentlichen Analysetätigkeit abgeschlossen sein, sodass während dieser Phase keine Änderungen vorgenommen werden können. Im Rahmen der Dokumentation ist es ebenfalls von zentraler Bedeutung, dass Zugriffe auf die zu analysierenden Datenbestände, als Teil der Beweismittelkette, protokolliert werden. Nachdem eine Transformation an den Bestandsdaten vorgenommen wurde, ist es notwendig, bereits bestehende Prüfsummen durch neu zu berechnende Prüfsummen zu ergänzen (Willer 2012, S. 86-94). Dieses Vorgehen ergibt sich aus der Anforderung der Wiederholbarkeit, wobei die Persistenz der Informationen unter Verwendung dieses Vorgehens sichergestellt werden kann.

Bezüglich der Anwendung der Visualisierungsmethoden und -werkzeuge sollten prozessbeteiligte Analysten und Ermittler geschult werden, sodass die ermittelten Nachweise in ein standardisiertes Berichtsformat überführt werden können. Diese Aspekte sollten während der Datenanalyse unter Verwendung von interaktiven Grafiken und anderen Formen der Daten- und Informationsvisualisierung beachtet werden, sodass die ermittelten Erkenntnisse auch im Rahmen der zivil- oder strafrechtlichen Bewertung eines Vorfalls genutzt werden können.

6.4 Ergebnisdokumentation

Interaktive Grafiken und statische Informationsgrafiken müssen eine zielgruppenorientierte Ereignisdokumentation unterstützen (Heer et al. 2005). Mit dieser Dokumentation wird der Verlauf eines Vorfalls beschrieben und der Betrachter wird in die Lage versetzt, das Vorfallsgeschehen rekonstruieren zu können. Hierzu müssen Hinweise, vorhandene Spuren und Ereignisse über eine zeitliche oder inhaltliche Beziehung verknüpft werden. Der Vorfall muss dazu auf Grundlage eines Verlaufprotokolls, in dem alle Einzelschritte der Ermittlungsphasen protokolliert sind, allgemein verständlich beschrieben werden. Mit dieser Beschreibung wird sichergestellt, dass die Analysetätigkeit bzw. jene Aktivitäten, die zum Erkenntnisgewinn geführt haben, wiederholbar sind. Ein optionales Ergebnisprotokoll kann zielgruppenorientiert gestaltet sein und beinhaltet einzelne Ereignisse im Vorfallsverlauf und deren

Abhängigkeit zu anderen vor- oder nachgelagerten Ereignissen. In diesem Schritt ist es möglich, Ereignisse und Objekte mit Personen in Verbindung zu bringen.

Die Manipulierbarkeit der Analyseergebnisse muss dabei ausgeschlossen werden. Sofern Zweifel an der Integrität der Ergebnisse oder des Verfahrens, das zu den Erkenntnissen geführt hat, bestehen, muss die Wiederholbarkeit der ausgeführten Schritte gegeben sein (BSI 2011, S. 22-25). Das setzt voraus, dass die verwendete Methode oder Anwendung die Grundanforderungen an die Wiederholbarkeit und Glaubhaftigkeit sowie die Beschreibung von Ursache und Wirkung erfüllen kann. Interaktive Grafiken und Informationsgrafiken, die während der Analyse erstellt wurden oder innerhalb der Dokumentation Verwendung finden, müssen daher so gespeichert werden, dass die Persistenz und Integrität der Daten sichergestellt ist. Neben der objektiven Darstellung des Vorfalls muss es möglich sein, die in der Dokumentation verwendeten Grafiken inhaltlich zu erweitern. Diese Aspekte sollten während der Datenanalyse unter Verwendung von interaktiven Grafiken und anderen Formen der Daten- und Informationsvisualisierung beachtet werden, sodass die ermittelten Erkenntnisse auch im Rahmen der zivil- oder strafrechtlichen Bewertung eines Vorfalls genutzt werden können.

7 Auswahlverfahren zur Bestimmung der Visualisierungsmethoden

Eine effektive Arbeitsweise während des forensischen Prozesses ermöglicht es, die Anzahl fallbezogener Anpassungen während der Vorfallsbearbeitung zu reduzieren. Neben einer normierten Dokumentation sollten daher auch die zu verwendenden Analysemethoden und -werkzeuge bestimmt werden. Dieses Vorgehen ermöglicht es, während der Untersuchung des Vorfalls aus bestehenden Methoden und Modellen die geeignetsten auszuwählen und die auszuführenden Tätigkeiten in der Analysephase auf diesen Hilfsmitteln aufzubauen.

Auch im Hinblick auf die Anforderungen, die an eine forensische Untersuchung und deren Nachweise gestellt werden, bietet dieses Vorgehen verschiedene Vorteile: So ist es möglich, die Dokumentation für eine Klasse von Vorfällen in Teilen generisch zu erstellen, da die zur Analyse verwendeten Hilfsmittel für diese Fallklasse aufgrund ihrer Standardisierung bereits bestimmt sind. Um bereits vor der Analysetätigkeit eine Aussage und schlussendlich auch eine Entscheidung hinsichtlich geeigneter Hilfsmittel treffen zu können, ist es notwendig, mögliche Vorfälle sowie deren Inhalt und Struktur zu klassifizieren. Ausgehend von dieser Klassifikation können unter Verwendung verschiedener Entscheidungsmodelle Methoden und Werkzeuge bestimmt werden, die während der Vorfallsanalyse eingesetzt werden sollen. Ziel dieser Entscheidung ist es, ein Verfahren zur Informationsvisualisierung zu bestimmen, das es ermöglicht, Ereignisse, die einen Vorfall beschreiben, wirksam und zielgerichtet zu analysieren.

7.1 Motivation

Die Erstellung eines Auswahlverfahrens, das es ermöglicht, die geeigneten Visualisierungsmethoden für die Darstellung von vorfallsrelevanten Informationen während der forensischen Datenanalyse zu bestimmen, unterstützt prozessbeteiligte Personen insofern, als dass sie die Möglichkeit erhalten, den Analyse- und Dokumentationsprozess weitestgehend zu standardisieren (Garfinkel 2010, S. 64-73). Die Grundlage hierfür bildet eine Auswahl, mit der die verfügbaren Darstellungsmethoden und -werkzeuge einzelnen Vorfallsklassen zugeordnet werden. Mögliche Einschränkungen in Bezug auf die Darstellungsmöglichkeiten, die sich aus der zu

analysierenden Datenstruktur ergeben, können so verhindert werden. Diese mehrstufige Auswahl ermöglicht es, eine Art der Visualisierung zu bestimmen, die im Kontext des Vorfalls die verfügbaren Informationen in einer allgemein verständlichen Form darstellt. Analysten und Ermittler werden auf diese Weise bei der Bestimmung von Nachweisen effektiv unterstützt und andere Nutzer können vorfallsrelevante Ereignisse auf Basis dieser Darstellung nachvollziehen (Osborne und Turnbull 2009, S. 1012-1017). Das Verfahren sollte sich dabei durch eine hohe Wirksamkeit und Bedienbarkeit auszeichnen.

Ziel ist es, dass die ermittelten Nachweise bereits während der Analysephase in einer für alle prozessbeteiligten Personen verständlichen Form aufbereitet werden können (Balakrishnan und Walker 2014). Dies ermöglicht den vereinfachten Zugang zu vorfallsrelevanten Ereignissen und deren Auswirkungen. Das Ziel hierbei ist die Abstraktion der Analysetätigkeit, sodass der Ermittlungsfokus auf der Darstellung von Ereignissen und den daraus resultierenden Vorfällen liegt. Im Hinblick auf die Dokumentation zeichnet sich dieses Vorgehen durch eine hohe Wirksamkeit aus. Der Grund hierfür ist die sorgfältige Prüfung, ob die zur Informationsdarstellung eingesetzten Methoden und Werkzeuge auch dazu geeignet sind. Da die Bestimmung der verwendeten Analysemethoden im Rahmen der Datenuntersuchung zu erfolgen hat, bietet es sich an, ein solches Auswahlverfahren in die Prozessphase zu übernehmen, da es hier stets möglich ist, das Verfahren zu erweitern oder anzupassen, sodass Anforderungen, die sich aus neuen oder geänderten Fallkonstellationen ergeben, berücksichtigt werden können.

7.2 Modellbildung

Um die Funktionen und das Zusammenwirken von Daten, Methoden und Werkzeugen zur Visualisierung von vorfallsrelevanten Daten und Informationen während einer forensischen Untersuchung zu beschreiben, wird im Rahmen dieser Arbeit ein erweiterbares Hilfsmittel entwickelt. Da die Summe aller Elemente nicht zwangsläufig die Gesamtheit eines Vorfalls ist, bietet eine vereinfachte, strukturgleiche Darstellung, in diesem Fall ein Vorgehensmodell, die Möglichkeit, bekannte Herausforderungen lösungsorientiert zu bearbeiten. Eine Einschränkung der Bearbeitungsmöglichkeiten im Hinblick auf eine Vorfallsart besteht nicht, da die in diesem Modell beschriebenen Bestandteile beliebig veränderbar sind. Der Aufbau des Vorgehensmodells wird in den

folgenden Abschnitten durch die Beschreibung der Modellgrundlagen, der Fall- und Ereignisklassifikation und der Bestimmung der Visualisierungsmethode erläutert.

7.2.1 Modellgrundlagen

Informationen und Daten, die einen Vorfall beschreiben, unterscheiden sich in Abhängigkeit von diesem Vorfall sowohl inhaltlich als auch strukturell. Um zu vermeiden, dass Methoden und Werkzeugen, die dazu geeignet sind, vorfallsrelevante Daten darzustellen, für jeden Vorfall individuell bestimmt werden müssen, ist es möglich, Vorfallsklassen sowie deren Inhalt und Struktur einmalig zu klassifizieren. Diese Klassifikation ermöglicht die standardisierte Auswahl von Analysemethoden. Bedeutende Faktoren sind hierbei das Volumen und der Strukturierungsgrad der Datenbestände.

In Bezug auf den Strukturierungsgrad lassen sich strukturierte, semistrukturierte und unstrukturierte Daten unterscheiden. Strukturierte Daten beschreiben die Information in einem definierten Format. Diese Struktur kann normalisiert und z. B. in einer relationalen Datenbank verarbeitet werden. Die Inhalte semistrukturierter Daten sind nicht zwangsläufig fest typisiert, sodass ihre Struktur beliebig erweitert werden kann. Ein Beispiel für semistrukturierte Daten sind Log-Daten und Tabellenkalkulationen. Es ist möglich, die Inhalte semistrukturierter Daten zu attribuieren, sodass ein entsprechendes Strukturmodell abgeleitet werden kann. Dieses ermöglicht die Interpretation und die Darstellung der Daten. Die Analyse von unstrukturierten Daten setzt ggf. eine Vorverarbeitung der Daten voraus, da diese im Regelfall keine formalisierte Struktur besitzen. Eine automatisierte Interpretation unstrukturierter Daten kann daher häufig nur unvollständig erfolgen. Beispiele für unstrukturierten Daten sind Bild- und Tonaufnahmen, Schriftsprache oder Archivdateien.

Um Fallklassen und deren Eigenschaften objektiv und nachhaltig zu beschreiben, ist es notwendig, das Format vorfallsrelevanter Datenbestände zu bestimmen und eine Einschätzung des Datenvolumens vorzunehmen. In der Bearbeitung von nutzerzentrierten Vorfällen, bei denen eine auf einem Endgerät ausgeführte Handlung den Kern des Vorfalls bildet, sind häufig populäre, in Teilen auch proprietäre, Dateiformate zu finden. Diese sind bei Vorfällen, bei denen das Verhalten einer zentralen

Netzwerkkomponente, z. B. über die Logdateianalyse, untersucht wird, ggf. nachgelagert von Bedeutung, bilden jedoch nicht den Kern der Untersuchungs- und Analysetätigkeit. Aus dieser Unterscheidung ergeben sich zentrale Erkenntnisse, die während der Vorfallsbearbeitung genutzt werden können, um die Analysetätigkeit zielgerichtet und damit auch effektiv zu gestalten.

7.2.2 Fall- und Ereignisklassifikation

Die Einsatzgebiete der IT-Forensik lassen sich unter verschiedenen Gesichtspunkten gliedern. Die folgende Übersicht kann daher beliebig gekürzt, erweitert oder verändert werden, sodass auch geänderte Anforderungen, die sich im Analyseverlauf ergeben, berücksichtigt werden können. Die hier beschriebene Fall- und Ereignisklassifikation beschreibt Straftaten und Arten von Vorfällen, die in einem informationstechnischen System oder durch ein solches begangen oder vorbereitet werden können (BKA 2003).

In den Tabellen 2 bis 6 werden je Element der Fall- und Ereignisklassifikation die folgenden Fragen betrachtet:

- Ereignisbeschreibung: Was kennzeichnet das Ereignis oder den Vorfall?
- Zeitraum: Mit welcher Häufigkeit oder über welchen Zeitraum tritt das Ereignis typischerweise auf?
- Datenvolumen: In welcher Höhe liegt typischerweise das Volumen an Quelldaten, die es während der Analyse zu berücksichtigen gilt?
- Dateinamenserweiterung: Welche Dateinamenserweiterung gilt es typischerweise bei der Datenanalyse zu berücksichtigen?

Die entsprechenden Antworten beschreiben Eigenschaften der einzelnen Vorfallsarten und bilden damit die Fall- und Ereignisklassifikation in der Domäne der Computerkriminalität.

Computerkriminalität - Erpressung	
Erpressung	Ereignisbeschreibung: § 253 Strafgesetzbuch (StGB) – Erpressung der Versuch, sich selbst oder Dritte durch die Androhung eines empfindlichen Übels oder Gewalt zu bereichern
	Zeitraum: isoliertes Ereignis, wiederkehrendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, einstelliger Gigabyte-Bereich
	Dateinamenserweiterung: .csv, .db, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .img, .iso, .jar, .java, .jpg, .kbm, .kdbx, .log, .m4p, .m4v, .md, .mdf, .mov, .mp3, .mp4, .mpeg, .msg, .odp, .ods, .odt, .onion, .pcab, .pdf, .pl, .png, .ppt, .pptx, .pst, .py, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .vbp, .wav, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 2: Fall- und Ereignisklassifikation – Computerkriminalität: Erpressung

Computerkriminalität - Spionage	
Spionage	Ereignisbeschreibung: die unrechtmäßige Beschaffung und das Erlangen von unbekannten Informationen oder geschütztem Wissen aus wirtschaftlichem oder politischem Interesse
	Zeitraum: isoliertes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Gigabyte-Bereich, einstelliger Terabyte-Bereich
	Dateinamenserweiterung: .aac, .bak, .bat, .bbl, .bfe, .c, .cbl, .cpp, .csv, .cvg, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .img, .iso, .jar, .java, .jpg, .kbm, .kdbx, .log, .m4p, .m4v, .md, .mdf, .mov, .mp3, .mp4, .mpeg, .msg, .odp, .ods, .odt, .onion, .ovm, .pcab, .pdf, .pl, .png, .ppk, .ppt, .pptx, .pst, .py, .r, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .vbox, .vbp, .wav, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 3: Fall- und Ereignisklassifikation – Computerkriminalität: Spionage

Computerkriminalität - Phishing	
Phishing	Ereignisbeschreibung: der Versuch, Daten und Informationen eines Benutzers unrechtmäßig zu erlangen, um diese zu missbrauchen (Identitätsdiebstahl)
	Zeitraum: isoliertes Ereignis, wiederkehrendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich
	Dateinamenserweiterung: .ahk, .csv, .doc, .docx, .exc, .exe, .gpg, .gz, .html, .jpg, .log, .msg, .odp, .ods, .odt, .pcab, .pdf, .png, .ppk, .ppt, .pptx, .pst, .rar, .svg, .swp, .tar, .tif, .txt, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 4: Fall- und Ereignisklassifikation – Computerkriminalität: Phishing

Computerkriminalität - Sabotage	
Sabotage	Ereignisbeschreibung: die vorsätzliche Störung eines Prozesses, z. B. durch die Beschädigung eines für die Durchführung notwendigen Objekts
	Zeitraum: isoliertes Ereignis, wiederkehrendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, mehrstelliger Terabyte-Bereich
	Dateinamenserweiterung: .ahk, .bat, .bbl, .bfe, .bib, .c, .cbl, .cpp, .csv, .cvg, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .img, .iso, .jar, .java, .jpg, .kbm, .kdbx, .log, .m4p, .m4v, .md, .mdf, .mov, .mp3, .mp4, .mpeg, .msg, .odp, .ods, .odt, .ovm, .pcab, .pdf, .pl, .png, .ppk, .ppt, .pptx, .pst, .py, .r, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .vbox, .vbp, .wav, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 5: Fall- und Ereignisklassifikation – Computerkriminalität: Sabotage

Computerkriminalität - Missbrauch	
Missbrauch	Ereignisbeschreibung: der den gesellschaftlichen oder rechtlichen Normen widersprechende Gebrauch von Mitteln und Gegenständen, z. B. zur Planung und Durchführung von Straftatbeständen
	Zeitraum: isoliertes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, einstelliger Terabyte-Bereich
	Dateinamenserweiterung: .ahk, .bak, .bat, .bbl, .bfe, .bib, .c, .cbl, .cpp, .csv, .cvg, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .img, .iso, .jar, .java, .jpg, .kbm, .kdbx, .log, .m4p, .m4v, .md, .mdf, .mov, .mp3, .mp4, .mpeg, .msg, .odp, .ods, .odt, .onion, .ovm, .pcab, .pdf, .pl, .png, .ppk, .ppt, .pptx, .pst, .py, .r, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .vbox, .vbp, .wav, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 6: Fall- und Ereignisklassifikation – Computerkriminalität: Missbrauch

Analog hierzu ist in den Tabellen 7 bis 11 die Fall- und Ereignisklassifikation für die Domäne der Wirtschaftskriminalität beschrieben.

Wirtschaftskriminalität - Geldwäsche	
Geldwäsche	Ereignisbeschreibung: Illegal erwirtschaftetes Geld oder ein illegal erworbener Vermögenswert wird durch einen Akteur in den legalen Finanz- und Wirtschaftskreislauf eingeschleust.
	Zeitraum: wiederkehrendes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Gigabyte-Bereich, einstelliger Terabyte-Bereich
	Dateinamenserweiterung: .aac, .bak, .bbl, .bfe, .bib, .cbl, .csv, .cvg, .db, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .img, .iso, .jar, .java, .jpg, .kbm, .kdbx, .log, .md, .mp3, .mpeg, .msg, .odp, .ods, .odt, .pdf, .png, .ppk, .ppt, .pptx, .pst, .py, .r, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .wav, .xls, .xlsx, .xml, .yml, .zip

Tabelle 7: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Geldwäsche

Wirtschaftskriminalität - Korruption	
Korruption	Ereignisbeschreibung: Missbrauch einer Vertrauensstellung durch Vorteilsannahme, auf die kein rechtmäßiger Anspruch besteht
	Zeitraum: isoliertes Ereignis, wiederkehrendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, einstelliger Gigabyte-Bereich
	Dateinamenserweiterung: .csv, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .ica, .jpg, .kbm, .kdbx, .log, .md, .mov, .mpeg, .msg, .odp, .ods, .odt, .pdf, .png, .ppk, .ppt, .pptx, .pst, .rar, .sav, .svg, .swp, .tar, .tif, .txt, .wav, .webm, .xls, .xlsx, .xml, .yml, .zip

Tabelle 8: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Korruption

Wirtschaftskriminalität - Insiderhandel	
Insiderhandel	Ereignisbeschreibung: die Verwendung von Insiderinformationen für Handelstransaktionen und Finanzmarktgeschäfte
	Zeitraum: isoliertes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, einstelliger Terabyte-Bereich
	Dateinamenserweiterung: .aac, .bak, .bbl, .bib, .c, .cbl, .csv, .db, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .jpg, .kdbx, .log, .mp3, .mpeg, .msg, .odp, .ods, .odt, .onion, .pdf, .png, .ppt, .pptx, .pst, .py, .r, .rar, .svg, .swp, .tar, .tif, .txt, .wav, .we, .xls, .xlsx, .xml, .yml, .zip

Tabelle 9: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Insiderhandel

Wirtschaftskriminalität - Insolvenzverfahren	
Insolvenzverfahren	Ereignisbeschreibung: ein Zustand, der durch akute oder drohende Zahlungsunfähigkeit gekennzeichnet ist und aus dem eine Überschuldung abgeleitet werden kann
	Zeitraum: isoliertes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Megabyte-Bereich, mehrstelliger Terabyte-Bereich
	Dateinamenserweiterung: .bak, .csb, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gpkg, .gz, .html, .img, .iso, .jar, .java, .jpg, .kdbx, .log, .msg, .odp, .ods, .odt, .ovm, .pcab, .pdf, .png, .ppk, .ppt, .pptx, .pst, .rar, .sav, .svg, .swp, .tar, .txt, .xls, .xlsx, .xml, .yml, .zip

Tabelle 10: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Insolvenzverfahren

Wirtschaftskriminalität - Marktmanipulation	
Marktmanipulation	Ereignisbeschreibung: Maßnahmen, die die Preisfindung des Markts rechtswidrig beeinflussen, verbunden mit dem Ziel, sich selbst oder Dritte zu bereichern oder Dritten zu schaden
	Zeitraum: isoliertes Ereignis, kurz- oder langanhaltendes Ereignis
	Datenvolumen: mehrstelliger Gigabyte-Bereich, einstelliger Terabyte-Bereich
	Dateinamenserweiterung: .bak, .bat, .cbl, .csv, .db, .dd, .doc, .docx, .exc, .exe, .gpg, .gz, .log, .msg, .odp, .ods, .odt, .pdf, .png, .ppt, .pptx, .pst, .py, .r, .rar, .svg, .swp, .tar, .tif, .txt, .xls, .xlsx, .xml, .yml, .zip

Tabelle 11: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Marktmanipulation

Zur Identifikation der Vorfallsklasse kann die Ereignisbeschreibung verwendet werden. Aus dieser lassen sich Informationen zum Ereigniszeitraum bestimmen, wobei der entsprechende Wert auf Grundlage ähnlicher Vorfälle berechnet und initial auch geschätzt werden kann. In Verbindung mit dem erwarteten Datenvolumen bieten diese Attribute einen Indikator zur Bestimmung eines möglichen Darstellungsverfahrens im Rahmen der Analysetätigkeit. Bezüglich der Datenvorverarbeitung und Transformation ist die Datenstruktur, die sich aus der Dateinamenserweiterung ableiten lässt, von zentraler Bedeutung. Eine Vorverarbeitung muss auch im Hinblick auf den angestrebten Standardisierungsgrad sichergestellt werden. Das Ziel ist eine hohe Datenqualität, sodass Abhängigkeiten, die sich im Verlauf der Visualisierung und der daran anknüpfenden Analyse ergeben, reduziert werden können. Die Zielsetzung der Vorverarbeitung und Transformation besteht in einem normierten und maschinenlesbaren Datenformat (Open Data Handbook 2019).

Auf Basis der dargestellten Vorfallsklassifikation ist es möglich, existierende Visualisierungsmethoden gemäß den Eigenschaften eines Vorfalls, insbesondere unter

Berücksichtigung der Datenstruktur, des Volumens und des Inhalts auszuwählen. Sofern sich Anpassungen innerhalb einer Domäne ergeben oder eine Domäne hinzugefügt wird, ist es notwendig, diese adäquat zu berücksichtigen. Dieses Vorgehen ermöglicht eine vereinfachte Unterscheidung der Fallarten und die Integration von neuen Datenquellen. Letzteres ist vor allem dann eine entscheidende Hilfestellung, wenn es nicht möglich ist, eine Fallart für den zu untersuchenden Vorfall zu bestimmen.

7.2.3 Bestimmung der Visualisierungsmethode

Die wirklichkeitsgetreue und adressatengerechte Aufbereitung von Informationen setzt voraus, dass bei der Auswahl der Darstellungsmöglichkeiten und Visualisierungsmethoden die Struktur, der Inhalt und der Umfang der relevanten Daten berücksichtigt werden. Die Darstellungsauswahl kann, und das gilt es zu berücksichtigen, unter verschiedenen Gesichtspunkten erfolgen. Im Rahmen dieser Arbeit werden die Darstellungsmöglichkeiten in fünf Funktionsklassen gegliedert. In Verbindung mit der Fall- und Ereignisklassifikation ermöglicht dieses Vorgehen eine effektive und damit zielgerichtete Auswahl der Visualisierungsmethode (Marty 2009, S. 161-237; Zoss 2019; Shneiderman 1996, S. 336-343).

1. Vergleich

Diese Form der Visualisierung ermöglicht es, im Hinblick auf eine oder mehrere Eigenschaften sowohl Ähnlichkeiten als auch Unterschiede zu bestimmen. Der Vergleichsparameter ist dabei grundsätzlich variabel und kann z. B. über die Anzahl der Elemente oder einen Zeitverlauf definiert werden. Die im Folgenden beschriebenen Visualisierungen bieten diese Möglichkeiten.

Mit einem Verlaufsdiagramm (siehe Abbildung 21) wird die Datenverteilung über ein kontinuierliches Intervall beschrieben. Dieses kann, wie im abgebildeten Beispiel, ein Zeitraum oder eine andere Bezugsgröße sein. Jeder Balken des Verlaufsdiagramms beschreibt ein Intervall. Der Nutzen für die Analysetätigkeit ergibt sich aus der dargestellten Werteverteilung.

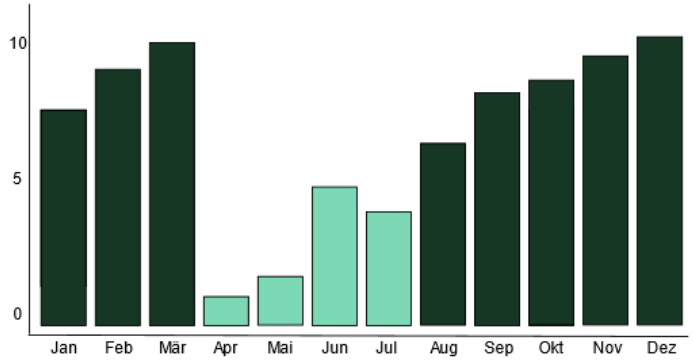

Abbildung 21: Visualisierungsmethode – Verlaufsdiagramm (Eigene Darstellung)

Mithilfe eines gruppierten Balkendiagramms (siehe Abbildung 22) werden zwei oder mehr Datenreihen visuell dargestellt. Ein Vergleich der Datenreihen erfolgt auf Grundlage zusammengefasster Kategorien. Die Länge der Balken wird dabei zum numerischen Vergleich der Datenreihen eingesetzt. Die Anzahl der Datenreihen beeinflusst die Lesbarkeit der Visualisierung und bildet damit einen wesentlichen Faktor in Bezug auf die Darstellungsgrenze der Visualisierungsmethode.

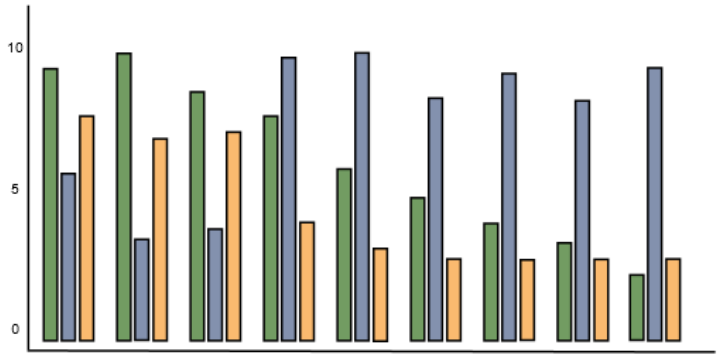

Abbildung 22: Visualisierungsmethode – gruppiertes Balkendiagramm (Eigene Darstellung)

Ein Blasendiagramm (siehe Abbildung 23) wird dazu verwendet, Beziehungen zwischen den auf der X- und Y-Achse abgetragenen Variablen darzustellen. Die Größe und die Proportion des Kreises, die durch ein Standardelement definiert wird, bilden eine dritte Variable und ermöglichen einen Vergleich mit anderen darzustellenden Elementen. Die Anzahl der Kreise beeinflusst die Lesbarkeit der Darstellung, sodass diese Anzahl und die Datengröße limitierende Faktoren bei der Erstellung der Visualisierung sind.

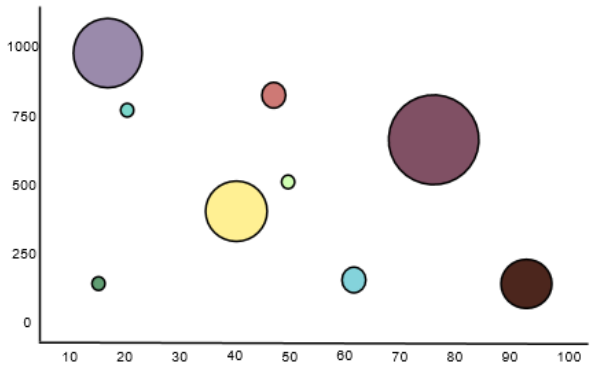

Abbildung 23: Visualisierungsmethode – Blasendiagramm (Eigene Darstellung)

2. Beziehung

Mit dieser Form der Visualisierung werden bestehende Verbindungen und Beziehungen zwischen verschiedenen Datensätzen durch die Korrelation einer oder mehrerer Variablen beschrieben (Shneiderman et al. 2011, S. 87-105). Die im Folgenden angeführten Visualisierungen bieten diese Möglichkeit.

In einem Matrix Diagramm (engl. heatmap; siehe Abbildung 24) sind Daten durch eine Variation der Farbgebung dargestellt. So ist es möglich, eine Kreuzuntersuchung durchzuführen, um Verbindungen über mehrere Variablen hinweg zu identifizieren. Eine Legende kann, in Abhängigkeit von der Datenkomplexität, die Wertebereiche voneinander abgrenzen. Das Matrix Diagramm kann auch dazu verwendet werden, Verbindungen in einem Zeitverlauf darzustellen.

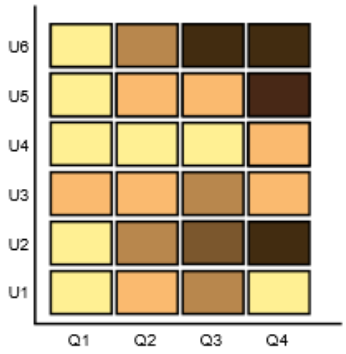

Abbildung 24: Visualisierungsmethode – Matrix-Diagramm (Eigene Darstellung)

Über ein Baumdiagramm (siehe Abbildung 25) können Beziehungen und Verbindungen in hierarchischer Form dargestellt werden. Der Ursprung des Diagramms ist in der Regel ein Wurzelknoten. Von diesem gehen Linienverbindungen aus, die eine Beziehung zu anderen Knoten beschreiben. Die Knoten können durch eine beliebige Variable besetzt werden.

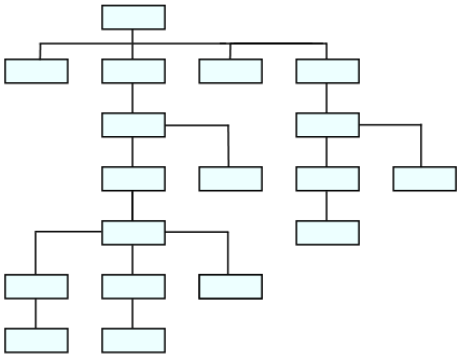

Abbildung 25: Visualisierungsmethode – Baumdiagramm (Eigene Darstellung)

Mithilfe eines Netzwerkdiagramms (siehe Abbildung 26) wird die Verbindung von einzelnen Datenelementen abgebildet, die wiederum durch Knoten dargestellt werden. Bestehende Beziehungen zwischen den Knoten werden durch den Einsatz von Verbindungslinien dargestellt. Die Strichstärke dieser Verbindungslinien kann in Abhängigkeit von der Variable, die die Beziehung beschreibt, variieren. Ein Netzwerkdiagramm ermöglicht es, einzelne Strukturen und deren Abhängigkeiten zu gruppieren. Die Lesbarkeit wird maßgeblich durch die Anzahl der darzustellenden Knoten bestimmt.

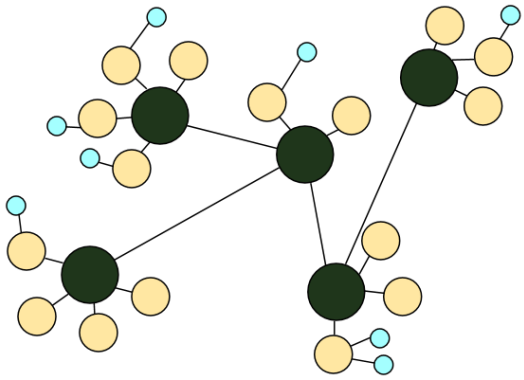

Abbildung 26: Visualisierungsbestimmung – Netzwerkdiagramm (Eigene Darstellung)

Mit einem Sehnendiagramm (siehe Abbildung 27) werden Zusammenhänge und Beziehungen zwischen verschiedenen Entitäten beschrieben. Eine Sehne stellt die Verbindung zwischen mehreren Attributen dar. Der Wert der Variable, die einer Beziehung zugeordnet ist, dargestellt durch die Sehnenstärke, wird dabei proportional berücksichtigt. Die Farbgebung ermöglicht die Darstellung mehrerer Attribute aus verschiedenen Kategorien.

Abbildung 27: Visualisierungsbestimmung – Sehnendiagramm (Eigene Darstellung)

3. Standort

Mit dieser Form der Visualisierung kann ein räumlicher, den Vorfall beschreibender, Bezug der Datensätze auf einer topografischen oder allgemein-geografischen Karte dargestellt werden. Die Informationen können durch einen zeitlichen Bezugspunkt versehen werden, um Veränderungen im Zeitverlauf darzustellen. Die im Folgenden genannten Visualisierungen bieten diese Möglichkeit.

Eine Punktverteilungskarte (siehe Abbildung 28) bietet die Möglichkeit, räumliche Muster oder die Verteilung von Daten über eine zu definierende Region dazustellen. Die Größe der Landmarken kann in Abhängigkeit von der darzustellenden Information gleich sein oder variieren. Die Aussagekraft hinsichtlich der Mustererkennung setzt eine Mindestanzahl von Elementen voraus.

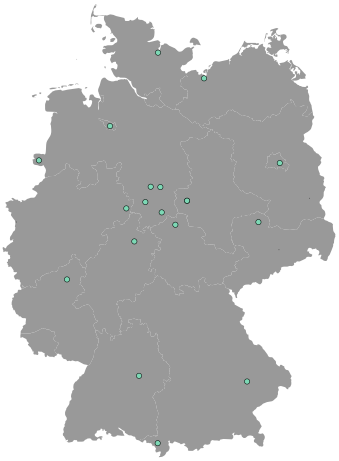

Abbildung 28: Visualisierungsmethode – Punktverteilungskarte (Eigene Darstellung)

Mit einer Verbindungskarte (siehe Abbildung 29) wird eine räumliche Verbindung zwischen zwei oder mehr Landmarken beschrieben, die durch eine oder mehrere Geraden verbunden werden. Die Stärke der Geraden kann dazu genutzt werden, die Intensität der Verbindung, z. B. bei Mittel- oder Güterbewegungen, darzustellen. Ein begrenzender Faktor hinsichtlich der Lesbarkeit ist die Dichte der darzustellenden Verbindungen, insbesondere in begrenzten Kartenausschnitten.

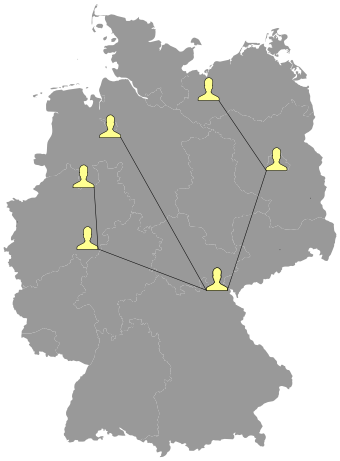

Abbildung 29: Visualisierungsmethode – Verbindungskarte (Eigene Darstellung)

In einer Blasenkarte (siehe Abbildung 30) wird eine geografische Region oder Landmarke durch einen Kreis markiert. Es ist möglich, dass dieser Kreis proportional der Größe der darzustellenden Variable entspricht oder dass der Variablenwert in die Darstellungsbeschriftung integriert ist. Die Grenzen dieser Visualisierungsmethode ergeben sich aus der Anzahl der darzustellenden Variablen und den daraus resultierenden Überlappungen.

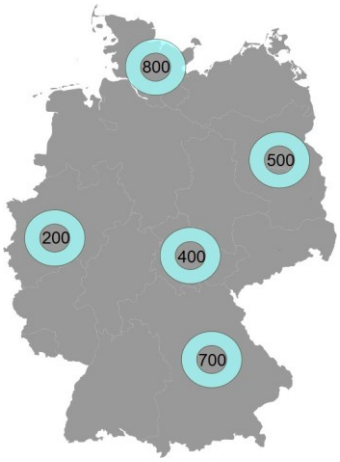

Abbildung 30: Visualisierungsmethode – Blasenkarte (Eigene Darstellung)

4. Verteilung

Diese Form der Visualisierung ermöglicht die Darstellung von Daten über ein beliebiges Intervall. Der Fokus liegt hierbei auf der Verteilung und der Häufigkeit eines Elements oder einer Gruppe von Elementen innerhalb einer Kategorie.

Mit einer Wortwolke (engl. tag cloud; siehe Abbildung 31) wird die Häufigkeit beschrieben, mit der ein Wort oder Wortstamm in einem beliebig großen Textkörper enthalten ist. Die Anzeigegröße ist abhängig von der Wortwiederholung. Durch den Ausschluss ausgewählter Wortarten, z. B. Artikel, Pronomen und Präpositionen, lassen sich Kontextinformationen zum Textkörper bestimmen. Diese Informationen können zur Erstellung von Such- und Stichwortlisten verwendet werden oder geben einen Hinweis auf den Sprachgebrauch einer sozialen Gruppe.

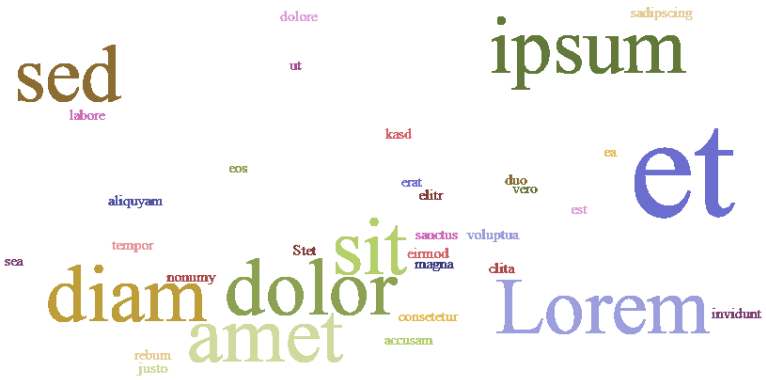

Abbildung 31: Visualisierungsmethode – Wortwolke (Eigene Darstellung)

In einem Dendrogramm (siehe Abbildung 32) wird eine hierarchische Gruppierung von Elementen in einer Baumstruktur dargestellt. Aus dieser Struktur lassen sich Informationen über die Elementverteilung innerhalb der Datenstruktur ableiten.

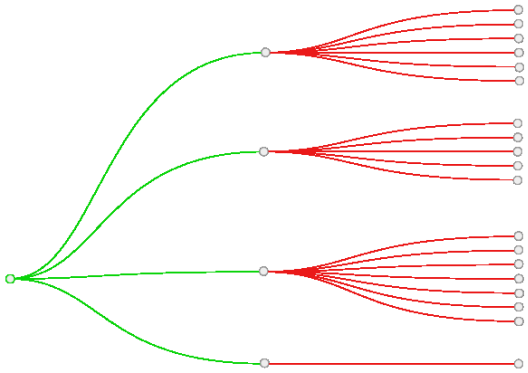

Abbildung 32: Visualisierungsmethode – Dendrogramm (Eigene Darstellung)

5. Zeitverlauf

Diese Form der Visualisierung kann dazu verwendet werden, Daten über einen frei wählbaren Zeitraum darzustellen. Der Nutzen liegt in der historisierten Darstellung von

Veränderungen. Diese Darstellungsweise ermöglicht es, Ereignisse zu ermitteln, die mit Anomalien und anderen Entwicklungen einhergehen. Zeitverläufe eigenen sich auch zum Verständnis und zur Dokumentation von isolierten oder wiederkehrenden Ereignissen.

Eine Zeitleiste (siehe Abbildung 33) bietet in diesem Zusammenhang die Möglichkeit, eine Liste von zusammenhängenden oder losen Ereignissen in chronologischer Reihenfolge darzustellen. Der Nutzen der Zeitleiste liegt in der Dokumentation und der Möglichkeit zur Analyse der darzustellenden Ereignisse. Die Zeitleiste kann somit auch als Ausgangspunkt für eine Vorfalls- und Ereignisbeschreibung verwendet werden.

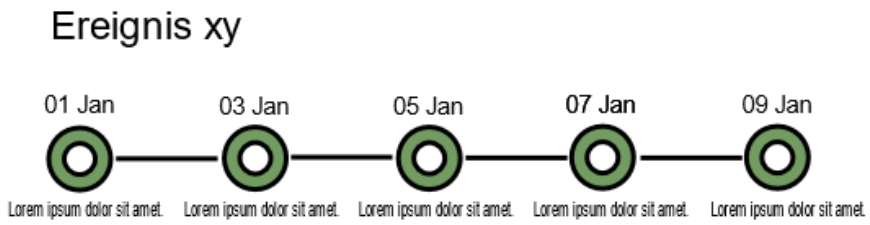

Abbildung 33: Visualisierungsmethode – Zeitleiste (Eigene Darstellung)

Ein Gantt-Diagramm (siehe Abbildung 34) kann dazu verwendet werden, durchgeführte und geplante Aktionen in chronologischer Reihenfolge darzustellen. Die Information über geplante und bereits durchgeführte Aktionen gibt Auskunft über den Ausführungsstatus der betroffenen Aktion und ermöglicht einen Ausblick auf Folgeaktivitäten.

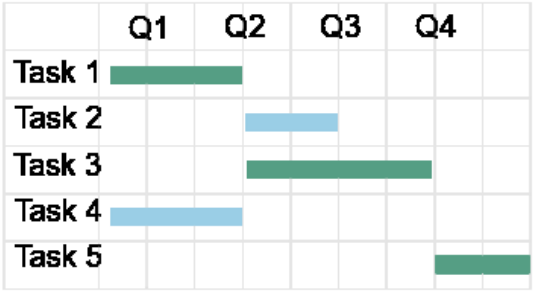

Abbildung 34: Visualisierungsmethode - Gantt-Diagramm (Eigene Darstellung)

Bei den in den fünf Funktionsklassen dargestellten Visualisierungsmethoden und Darstellungsmöglichkeiten kann es sich um interaktive Grafiken und statische Informationsgrafiken handeln. Dem Anwender oder Ersteller der interaktiven Grafik ist es stets möglich, den Darstellungsinhalt zu verändern. Elemente können somit aggregiert,

ausgeschlossen oder selektiert werden. Der Analyst oder Ermittler wird dadurch in die Lage versetzt, stets und frei von Einschränkungen über den Darstellungsinhalt zu verfügen. Diese Eigenschaft unterscheidet die interaktive Grafik von der statischen Informationsgrafik. Mögliche Schnittmengen der beschriebenen Funktionsklassen ergeben sich durch die Deutungsfreiheit bzgl. der einzelnen Visualisierungsmethoden und Darstellungsmöglichkeiten sowie deren Verbreitung. Der Intention einer inhaltsorientierten Darstellung von Vorfallsdaten muss dabei stets Rechnung getragen werden, denn der Begriff ‚Visualisierung' deutet vor allem darauf hin, dass Daten und Informationen in eine sichtbare Form überführt werden sollen.

7.3 Auswahlverfahren

Auf Grundlage der Fall- und Ereignisklassifikation ist es in Verbindung mit den verfügbaren Visualisierungsmethoden und Darstellungsformen möglich, ein Auswahlverfahren zur fall- und ereignisorientierten Bestimmung der geeigneten Visualisierung zu erstellen. Die Auswahlmöglichkeiten innerhalb dieses Modells sind auf eine Menge von Elementen begrenzt, die den Anforderungen und Eigenschaften der Fallklassifikation entsprechen. Durch den Rückschluss, der sich aus den Eigenschaften und Nutzungsmöglichkeiten der zur Verfügung stehenden Grundgesamtheit (siehe Kapitel 7.2.3) ergibt, ist es möglich, die Eignung eines Darstellungsverfahrens oder einer Methode im Hinblick auf die Darstellung einzelner Ereignisse oder Vorfälle zu bestimmen. Die darzustellenden Vorfälle oder Ereignisse weisen bestimmbare Eigenschaften auf. Diese ermöglichen es, einen direkten Vergleich zwischen den bekannten Eigenschaften der Vorfälle sowie deren inhaltlicher Struktur und den verfügbaren Darstellungsmethoden vorzunehmen (Varga et al. 2013). Eine ausgeprägte Übereinstimmung der Vergleichsparameter verweist auf die Eignung des Verfahrens dafür, die vorhandenen Ereignisse des Vorfalls zielgruppengerecht darzustellen. Hierfür gibt es je Vorfall, der in der Fall- und Ereignisklassifikation beschrieben wird, eine Informationsgrafik, die im Hinblick auf die Ereignisbeschreibung, den Zeitraum und das Datenvolumen eine Darstellungsempfehlung anbietet. Die Aufteilung erfolgt analog zur Fall- und Ereignisklassifikation in den Domänen Computerkriminalität (mit den Ausprägungen Erpressung, Spionage, Phishing, Sabotage und Missbrauch) und

Wirtschaftskriminalität (mit den Ausprägungen Geldwäsche, Korruption, Insiderhandel, Insolvenzverfahren und Marktmanipulation).

In den folgenden Informationsgrafiken werden Empfehlungen hinsichtlich der Darstellungsmöglichkeiten und Visualisierungsmethoden zur Analyse verschiedener Vorfallsarten aufgezeigt. Diese können von einem Anwender zur Beantwortung mehrerer W-Fragestellungen innerhalb des forensischen Prozesses genutzt werden. Interaktive Grafiken zur Beschreibung vorfallsrelevanter Daten können im Anschluss an die Analysetätigkeit auch zur Dokumentation der ermittelten Erkenntnisse sowie zur Einordnung der Ermittlungsergebnisse nach der CERT-Taxonomie eingesetzt werden (BSI 2011, S. 29-38).

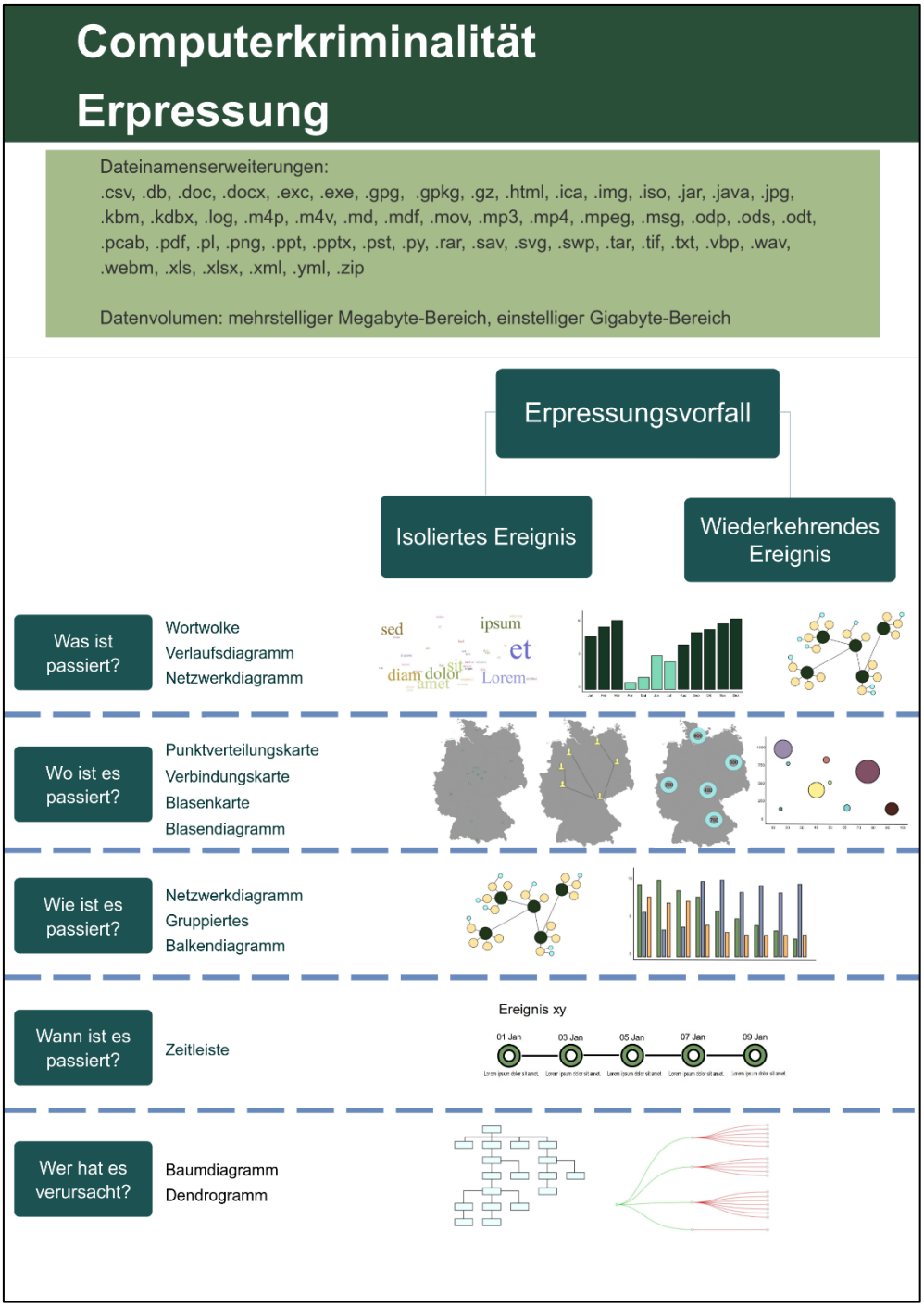

Abbildung 35: Visualisierungsauswahl für die Fallklasse ‚Erpressung' aus der Domäne Computerkriminalität (Eigene Darstellung)

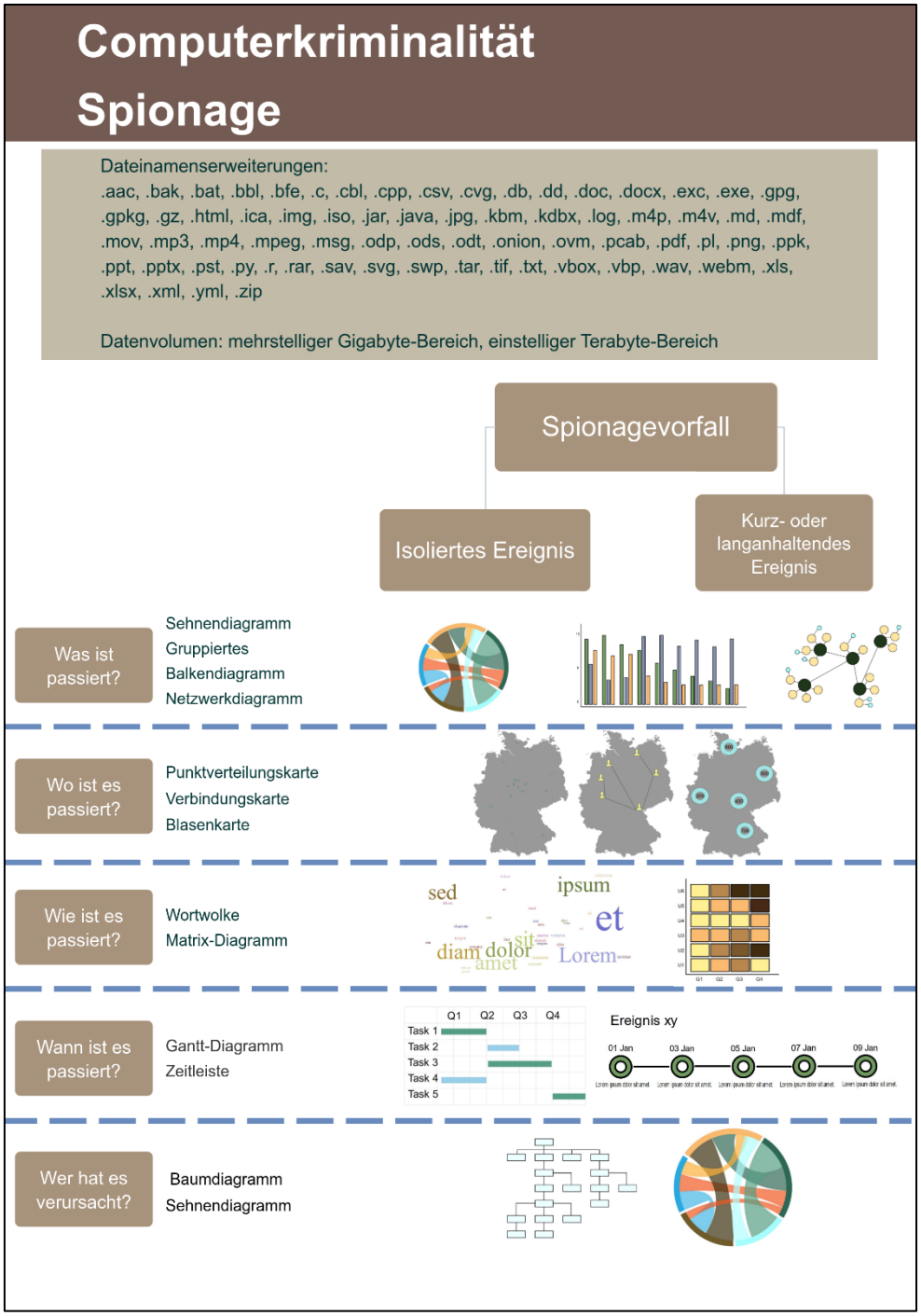

Abbildung 36: Visualisierungsauswahl für die Fallklasse ‚Spionage' aus der Domäne Computerkriminalität (Eigene Darstellung)

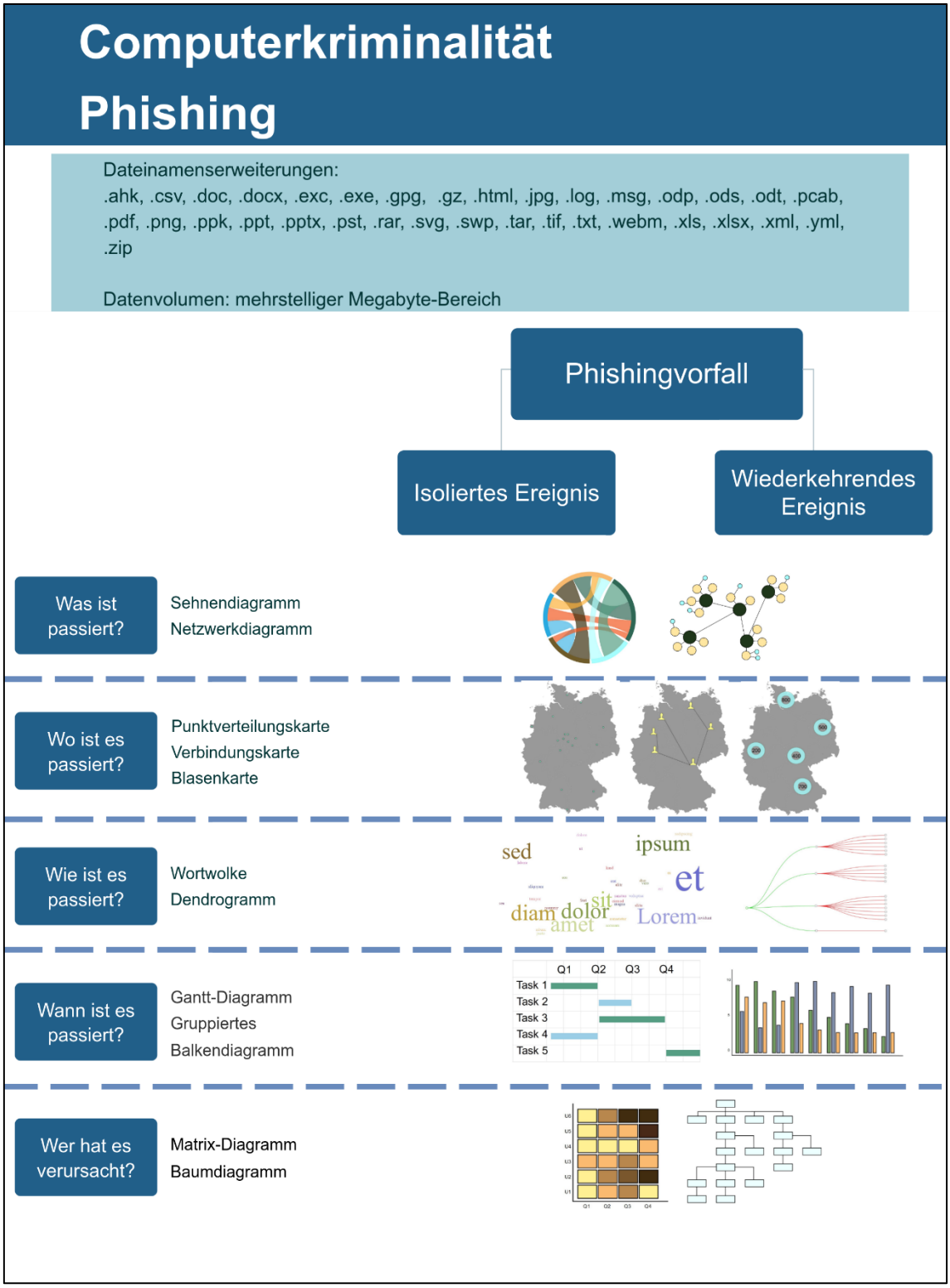

Abbildung 37: Visualisierungsauswahl für die Fallklasse ‚Phishing' aus der Domäne Computerkriminalität (Eigene Darstellung)

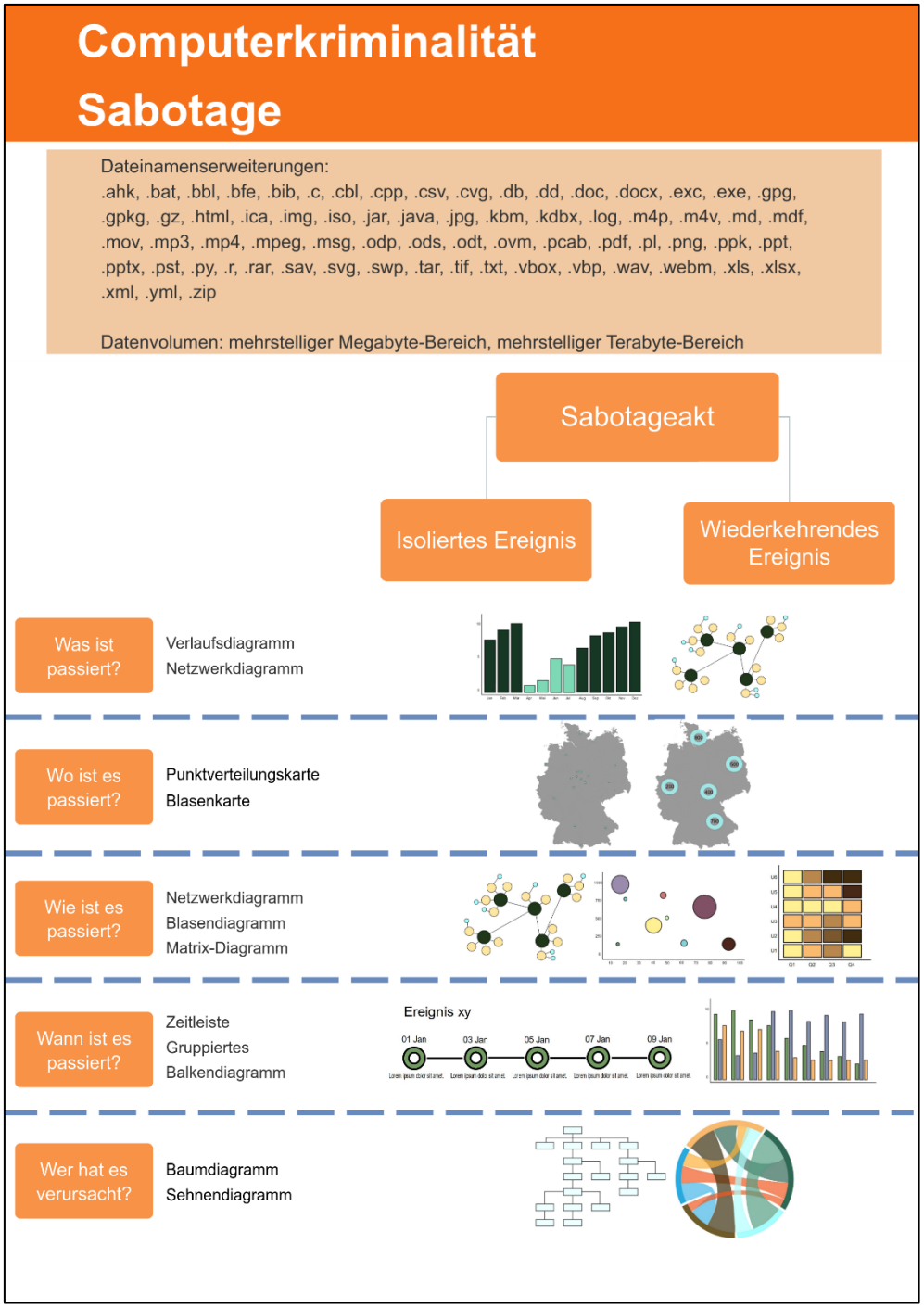

Abbildung 38: Visualisierungsauswahl für die Fallklasse ‚Sabotage' aus der Domäne Computerkriminalität (Eigene Darstellung)

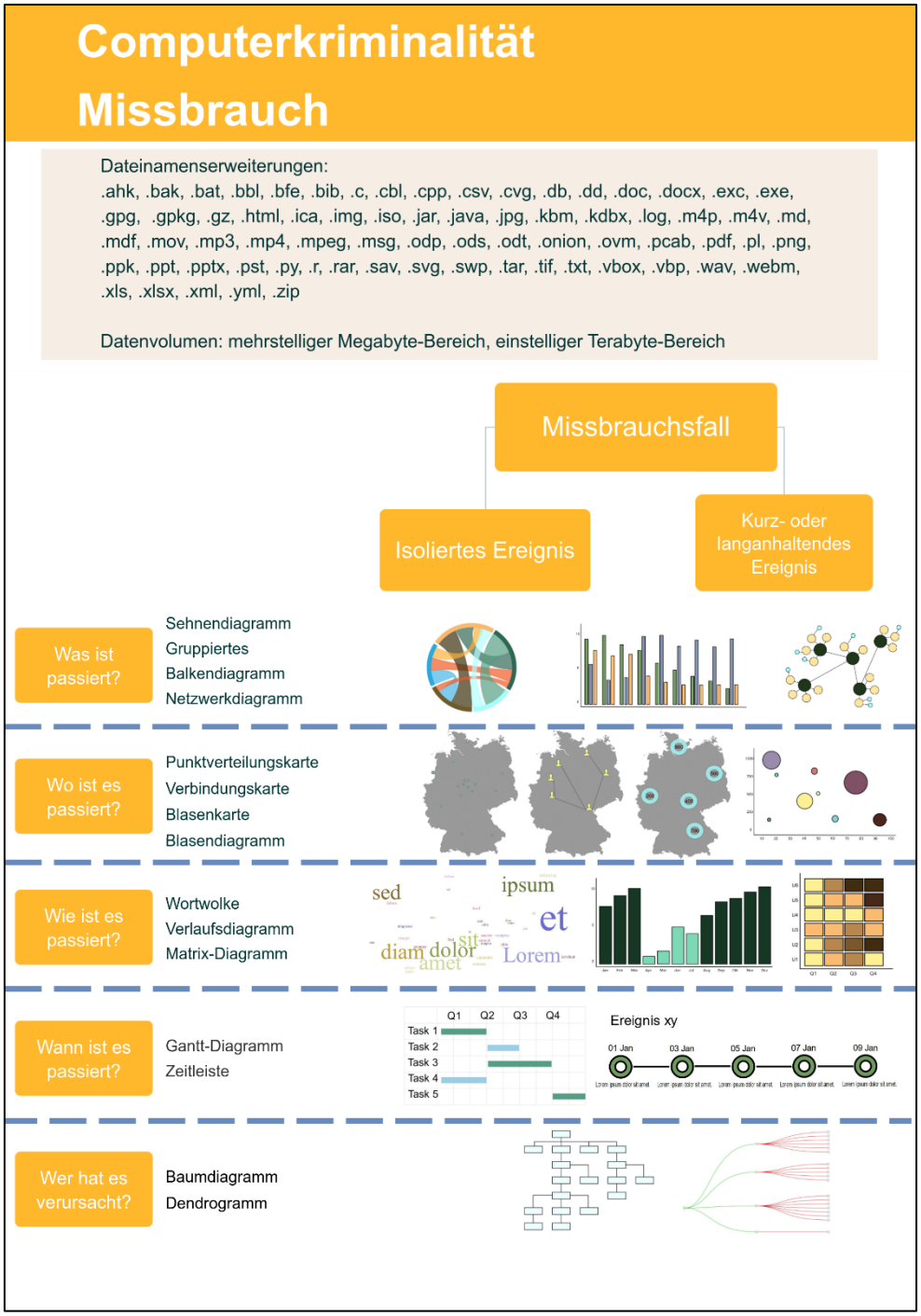

Abbildung 39: Visualisierungsauswahl für die Fallklasse ‚Missbrauch' aus der Domäne Computerkriminalität (Eigene Darstellung)

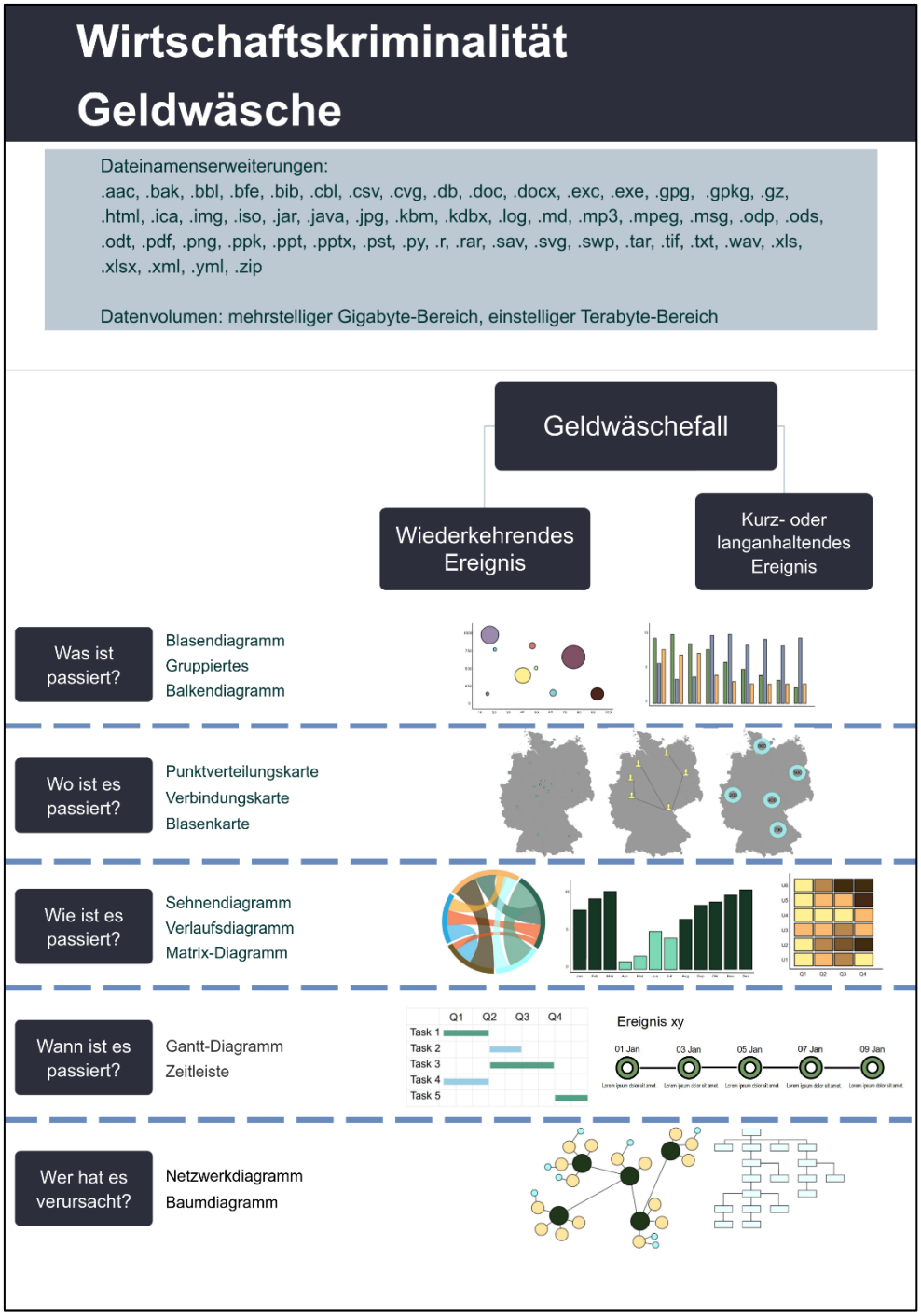

Abbildung 40: Visualisierungsauswahl für die Fallklasse ‚Geldwäsche' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung)

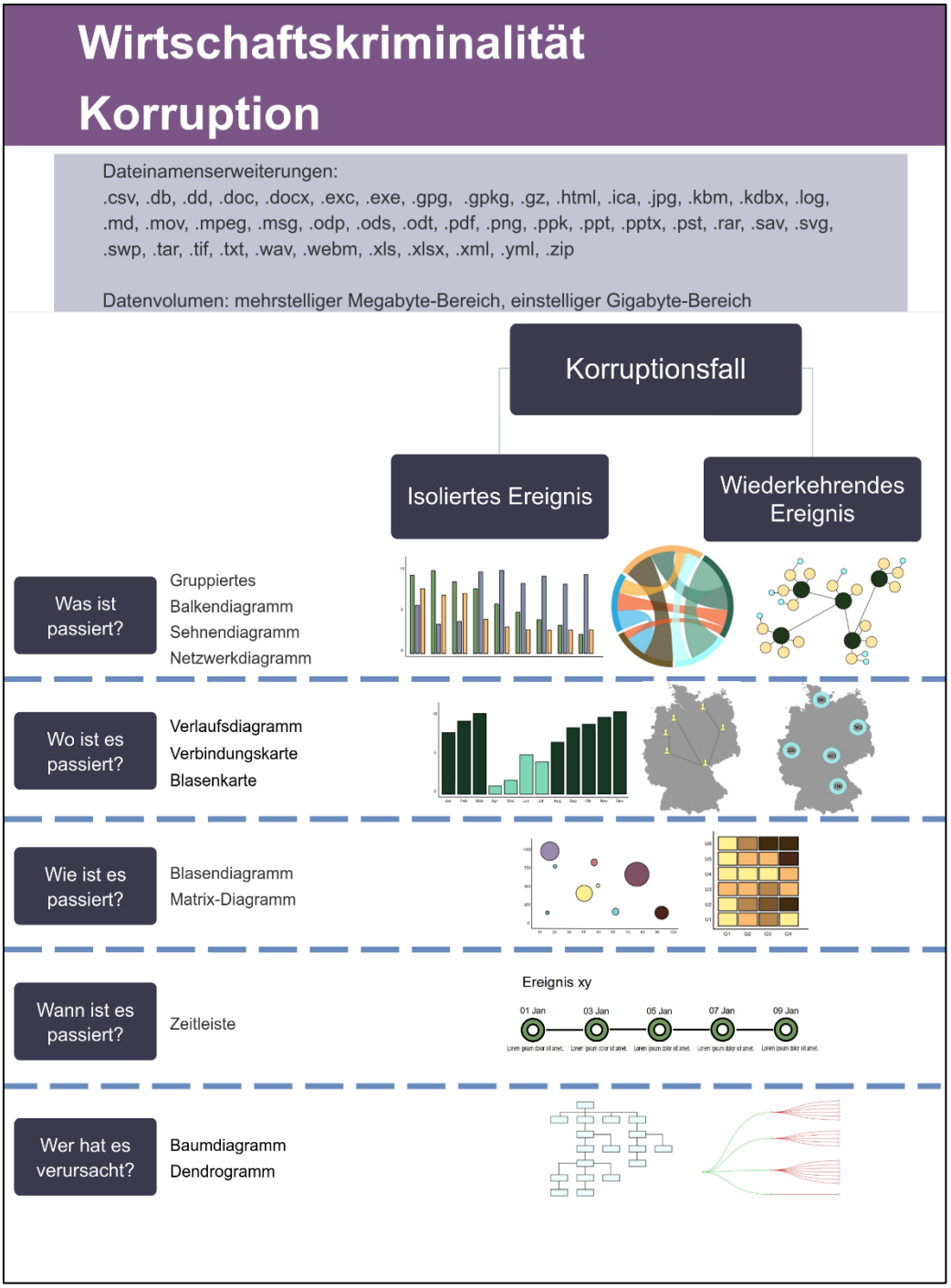

Abbildung 41: Visualisierungsauswahl für die Fallklasse ‚Korruption' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung)

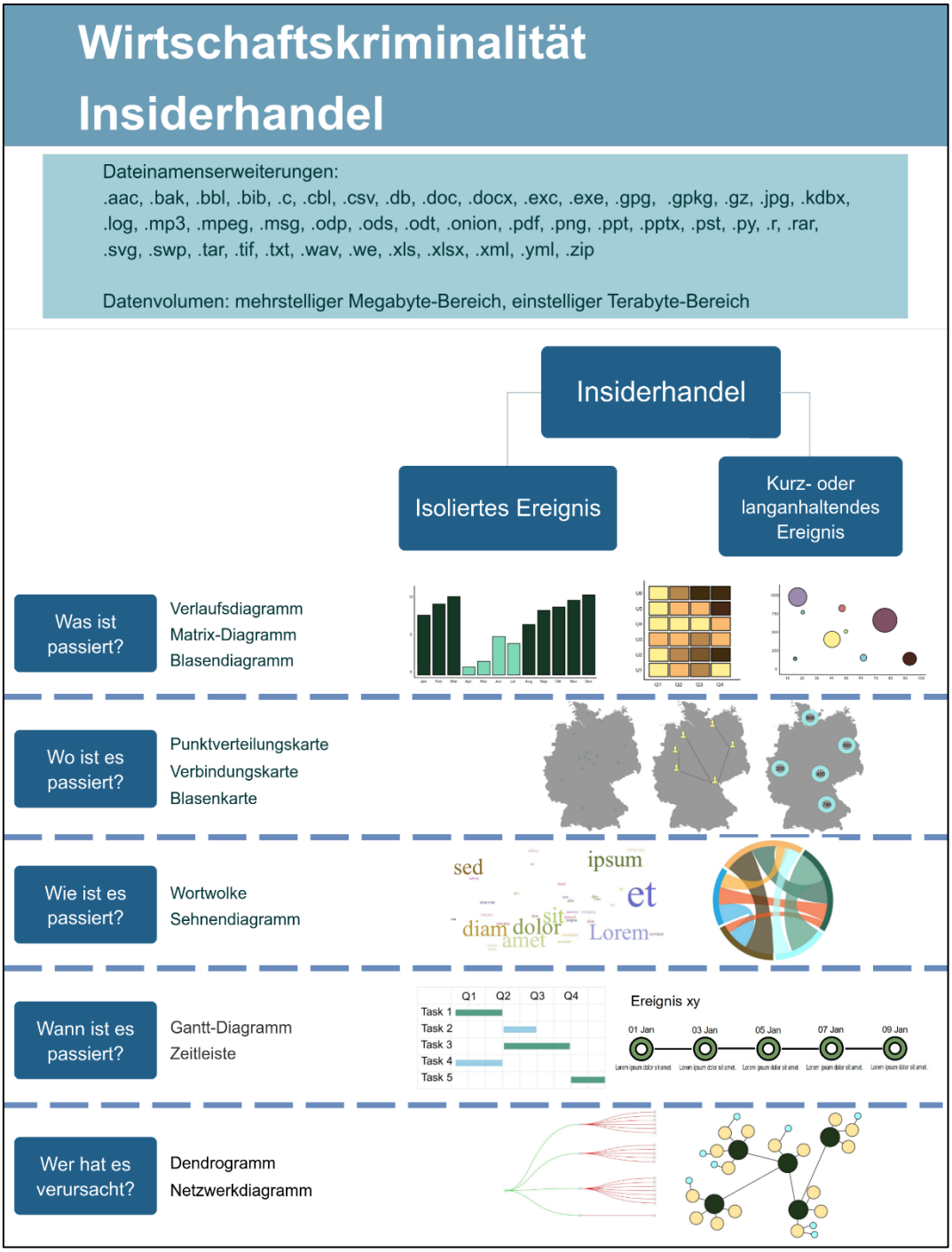

Abbildung 42: Visualisierungsauswahl für die Fallklasse ‚Insiderhandel' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung)

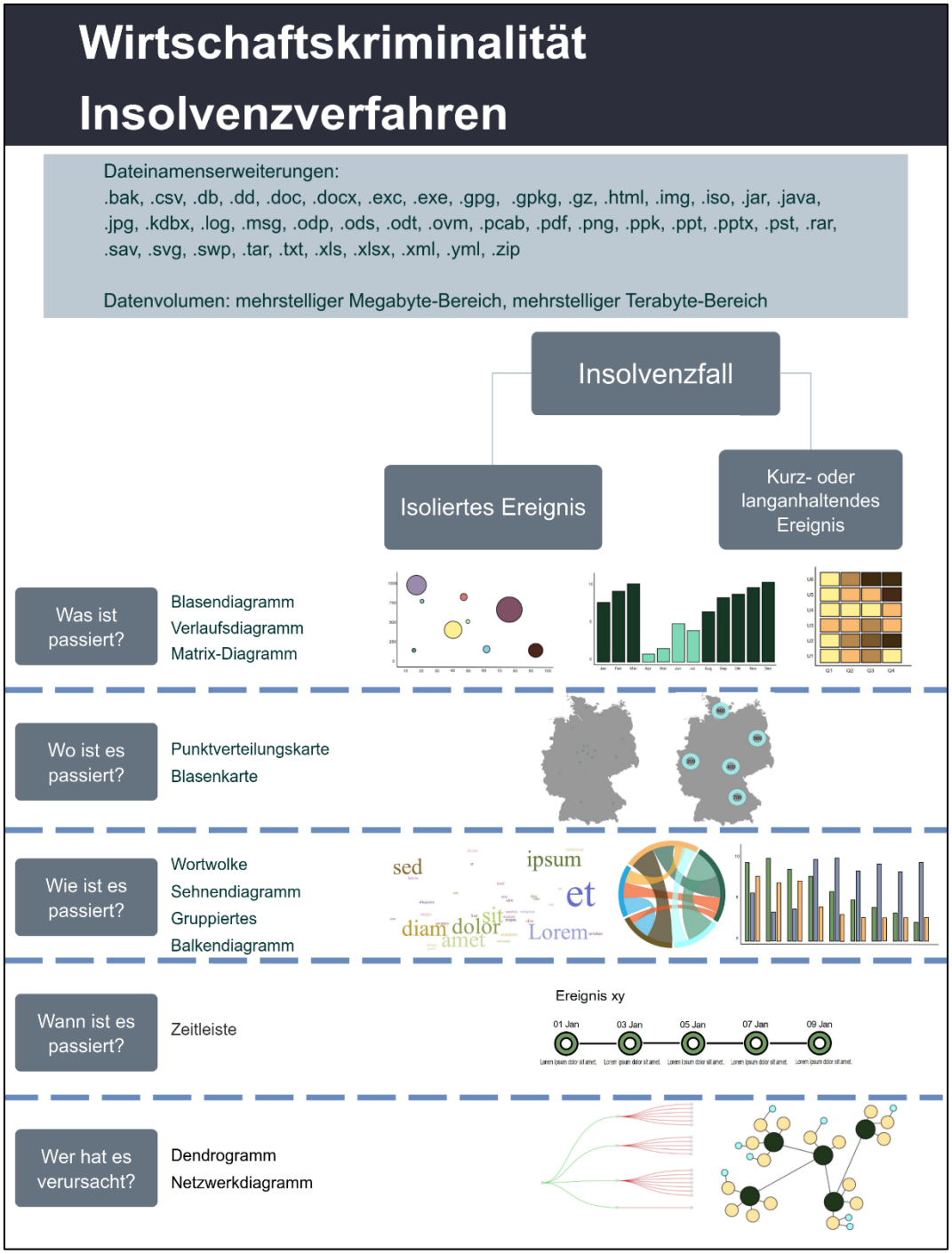

Abbildung 43: Visualisierungsauswahl für die Fallklasse ‚Insolvenzverfahren' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung)

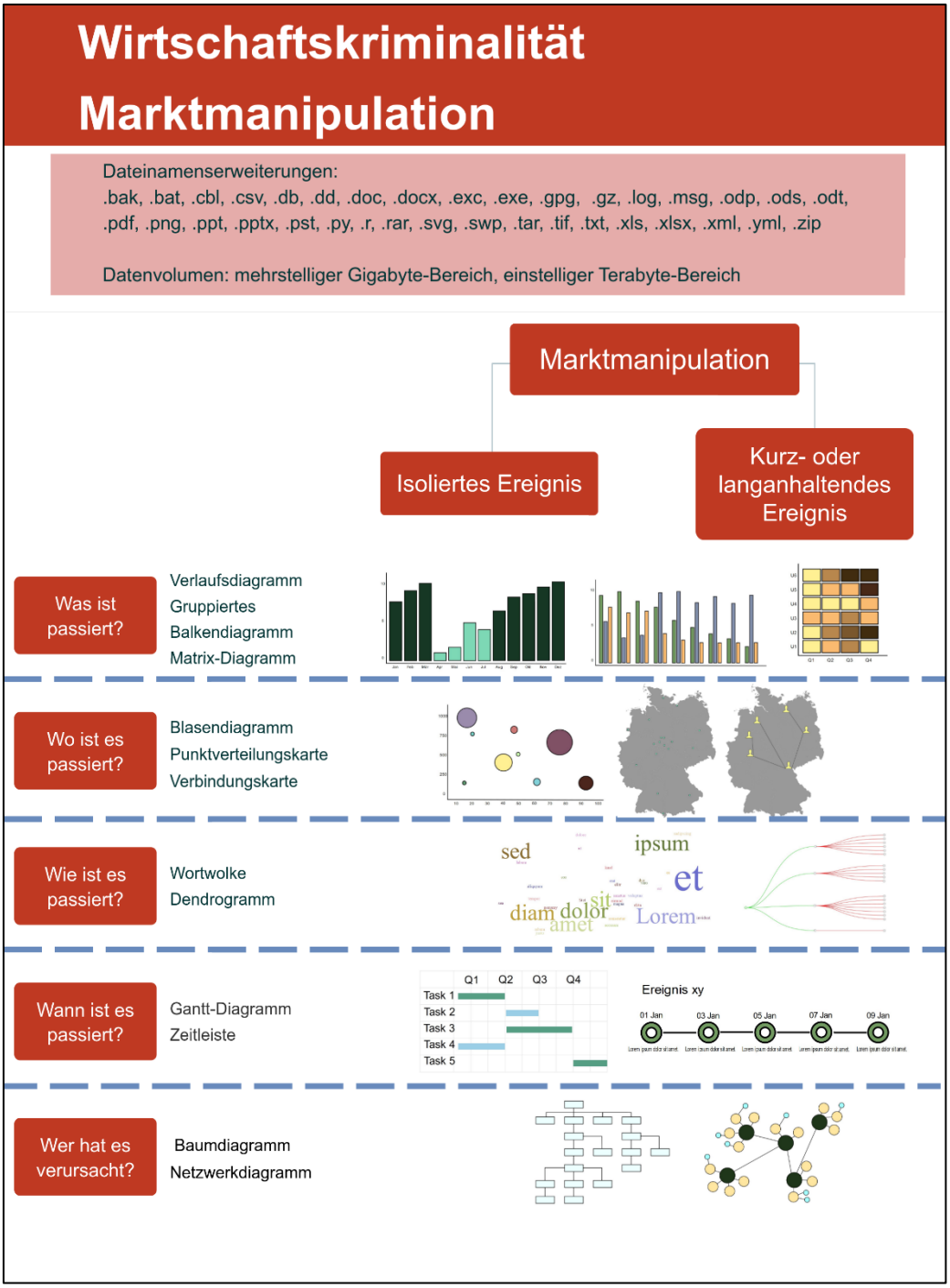

Abbildung 44: Visualisierungsauswahl für die Fallklasse ‚Marktmanipulation' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung)

Diese Empfehlungen zur Darstellung vorfallsrelevanter Daten in den Domänen der Computer- und Wirtschaftskriminalität unterstützen Analysten und Ermittler in mehreren Phasen des forensischen Prozesses. Nutzungsmöglichkeiten bestehen insbesondere in den Phasen der Datenuntersuchung, der Analyse und der Dokumentation. Die Auswahl einer geeigneten Visualisierungsmöglichkeit zur adressatengerechten Darstellung eines

Vorfalls stellt indes noch keine Analyseplattform dar, die die Erstellung unterstützt oder ermöglicht.

Ein alternatives Vorgehen zur Datenanalyse mithilfe interaktiver Grafiken könnte die initiale Bestimmung einer Analyseplattform zum Ziel haben. Die Visualisierungsbestimmung wäre infolgedessen eine Tätigkeit, die im Anschluss an die Auswahl der Analyseplattform ausgeführt würde. Die Auswahl der Visualisierungsmöglichkeiten ist dann auf die auf dieser Plattform verfügbaren Möglichkeiten beschränkt. Des Weiteren unterliegt die Bestimmung einer Analyseplattform weitreichenden Abhängigkeiten, angefangen bei der Mittelverfügbarkeit, die sich aus dem Spannungsfeld hinsichtlich der verfügbaren Zeit-, Kosten- und Personalressourcen ergibt, bis hin zu regulatorischen Anforderungen zwecks der Datenverarbeitung in Rechnernetzen (Public oder Community-Cloud), Import- und Export-Embargos oder gesetzlichen Einschränkungen hinsichtlich des Kundenstamms einer entsprechenden Anwendungs- oder Systemkomponente.

Vor diesem Hintergrund ist die von der Visualisierungsbestimmung ausgehende Plattformauswahl der allgemeingültige und zugleich lösungsorientierte Weg zur forensischen Datenanalyse unter Verwendung interaktiver Grafiken. Dennoch ist es wichtig, zu betonen, dass die Deutungshoheit hinsichtlich der Eignungsbewertung einer Darstellungsform beim Ersteller einer Visualisierung liegt. Eine abschließende ggf. einzelfallorientierte Bewertung bzgl. der Eignung einer Visualisierung zur Darstellung vorfallsrelevanter Informationen kann infolgedessen auch von den Empfehlungen für dieses Auswahlverfahren abweichen.

8 Anwendung des Auswahlverfahrens zur Visualisierungserstellung

Ziel dieses Abschnitts ist die Erprobung der in Kapitel 7 vorgestellten Visualisierungsauswahl anhand eines praxisnahen Fallbeispiels. Der Fokus bei der Erprobung des Verfahrens liegt auf der Unterstützung einer zielgerichteten inhaltsorientierten Analysetätigkeit. Vor diesem Hintergrund erfolgt im Anschluss an die Fallbestimmung die Auswahl einer Visualisierung für das Fallbeispiel. Um die Eignung der Darstellungsmöglichkeiten zu validieren, werden ausgewählte Visualisierungsmethoden unter Verwendung einer Analyseplattform implementiert. Nach Abschluss der Implementierung werden die Darstellungen und deren Nutzen im Rahmen der forensischen Datenanalyse bewertet.

8.1 Fallbestimmung und Datenstruktur

Die Bestimmung eines Testfalls, der ein realitätsnahes Ereignis auf Basis von Testdaten darstellt und sich somit auch für die Durchführung forensischer Datenanalysen unter Verwendung von interaktiven Grafiken eignet, erfolgte über die Website ‚DigitalCorpora.org'. Der Inhalt dieser Website umfasst neben Blog- und Foreneinträgen auch frei verfügbare Speicherabbilder (Haupt- und Festspeicher), Netzwerkpaketaufzeichnungen und Szenarien für Test- und Trainingszwecke. Das aktuellste der hier veröffentlichten Szenarien trägt den Titel ‚2018 Lone Wolf Scenario' (Digital Corpora 2018). Es wurde im Jahr 2018 von Thomas Moore erstellt. Vor diesem Hintergrund ist es nachvollziehbar, dass die Anzahl der verfügbaren Publikationen, die diesen Datensatz adressieren, beschränkt ist. Die Begründung hinsichtlich der Auswahl dieses Szenarios besteht neben der Aktualität auch in der Tatsache, dass es sich bei diesem Testfall um ein vollständiges Datenträgerabbild handelt.

Inhaltlich beschreibt dieser erdachte Fall die Planung einer Massenerschießung ausgehend von einer fiktiven Person, die in Folge einer polizeilichen Ermittlung verhaftet wurde. Den Auslöser für die Ermittlung bildete ein Hinweis, der von dem Bruder des Beschuldigten übermittelt wurde. Die Aufgabe des Ermittlers besteht darin, herauszufinden, ob es sich um einen Einzeltäter oder eine Gruppe von Tätern handelt. Die gesamte, auf einem mobilen Endgerät ausgeführte Planungstätigkeit und die damit

verbundenen Spuren und Nachweise innerhalb des Dateisystems wurden auf Grundlage dieser Tatsache mithilfe der Anwendung FTK Imager in Form eines Datenträgerabbilds gesichert (AccessData Forensic Toolkit (FTK) 2017). Dieses Speicherabbild wurde in Verbindung mit den Erkenntnissen aus einer vorgelagerten Datenuntersuchung veröffentlicht. Die während der Datenuntersuchung verwendeten Anwendungen sind Autopsy und Magnet Axiom.

Auf Grundlage dieser Informationen werden in Kapitel 8.2 die Darstellungsmöglichkeiten und Visualisierungsmethoden zur grafischen Darstellung der einzelnen Ereignisse des Vorfalls bestimmt. Diese Bestimmung erfolgt über die Fall- und Ereignisklassifikation, da diese den Startpunkt des Auswahlverfahrens bildet.

8.2 Implementierung des Fallbeispiels

Aus der Beschreibung des ausgewählten Szenarios geht hervor, dass die im Fallbeispiel beschuldigte Person sich elektronischer Hilfsmittel bedient hat, um eine strafbare Handlung zu planen. Unter Berücksichtigung des Auswahlverfahrens zur Visualisierungserstellung kann diese Handlung als Missbrauch gewertet werden. Infolgedessen kann die Infografik ‚Missbrauch' der Domäne Computerkriminalität dazu verwendet werden, Darstellungsmöglichkeiten zu bestimmen, mit denen vorfallsrelevante Daten und Informationen in Form von interaktiven Grafiken abgebildet werden können. Um die Informationen aus der vorgelagerten Datenuntersuchung zu analysieren und diese in eine interaktive Grafik bzw. mehrere interaktive Grafiken zu überführen, werden die Daten zuerst eingelesen, normalisiert und im Anschluss daran verarbeitet. Im Rahmen der Datenuntersuchung wurden vierzehn Dateisammlungen exportiert. Der initiale Import dieser Dateisammlungen, sodass diese in der browserbasierten Open-Source-Analyseplattform Kibana analysiert werden können, erfolgt über Logstash. Logstash ist eine serverseitige Open-Source-Anwendung, die in diesem Kontext zur Verarbeitung von Daten aus unterschiedlichen Quellen verwendet wird (Logstash 2019). Im Zuge dieser Ersterfassung werden einzelne Feldbezeichnungen und deren Inhalte normalisiert. Die Normalisierung umfasst ausschließlich beschreibende Kontextinformationen wie Zeit- und Datumsangaben. Daten, die den Vorfall inhaltlich beschreiben, bleiben unverändert, d. h. sie werden weder normalisiert noch transformiert.

Die Weiterleitung der normalisierten Datensammlungen erfolgt ausgehend von Logstash in eine Elasticsearch-Instanz. In dieser Instanz werden die erhaltenen Dokumente und Daten in einem NoSQL-Format gespeichert. Elasticsearch ermöglicht auf Basis von Lucene die Volltextsuche in den verwalteten Datenbeständen (Elasticsearch 2019). Auf der Analyseplattform Kibana kann diese Suchfunktion dazu verwendet werden, relevante Inhalte zu selektieren und damit den Umfang und die Struktur der Darstellung zu beeinflussen (Kibana 2019).

Abbildung 45 enthält die Programmanweisung zur Erfassung, Verarbeitung und Normalisierung der Datensammlung ‚Word Documents'. Die in der Datensammlung enthaltenen Daten werden durch Logstash erfasst und in Bezug auf die Zeit- und Datumsinformationen normalisiert. Im Anschluss an diese Verarbeitung werden die Daten in der Anwendung Elasticsearch in den Index ‚word_files' überführt.

```
input
  {
  file{
    start_position => "beginning"
    path => "C:/Users/User/Desktop/case_evidence_MA/Word Documents.csv"
    }
  }
filter
  {
  csv{
    separator => ","
    columns =>
    ["Record","Tags","Comments","Filename",
    "File","File System Last Modified Date/Time - UTC+00:00 (M/d/yyyy)",
    "File System Last Accessed Date/Time - UTC+00:00 (M/d/yyyy)",
    "File System Created Date/Time - UTC+00:00 (M/d/yyyy)",
    "Size (Bytes)","Title","Saved Size (Bytes)","Subject","Authors","Keywords","Comments",
    "Last Author", "Last Printed Date/Time - UTC+00:00 (M/d/yyyy)",
    "Last Modified Date/Time - UTC+00:00 (M/d/yyyy)", "Created Date/Time - UTC+00:00 (M/d/yyyy)",
    "Company","Source","Location","Evidence number"]
      }

    date{
    match => ["File System Last Modified Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "File System Last Modified Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

    date{
    match => ["File System Last Accessed Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "File System Last Accessed Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

    date{
    match => ["File System Created Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "File System Created Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

    date{
    match => ["Last Printed Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "Last Printed Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

    date{
    match => ["Last Modified Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "Last Modified Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

    date{
    match => ["Created Date/Time - UTC+00:00 (M/d/yyyy)", "M/d/yyyy h:mm:ss a"]
    target => "Created Date/Time - UTC+00:00 (M/d/yyyy)"
    timezone => "America/Chicago"
      }

  }
output{
  elasticsearch {
    hosts => ["localhost:9200"]
    index => "word_files"
  #document_type => "_doc"
        }
}
        stdout { codec => "dots" }
        if "_dateparsefailure" in [tags] {
                        stdout{codec => "rubydebug"}
```

Abbildung 45: Programmanweisung zur Erfassung, Verarbeitung und Normalisierung des Index ‚Word Documents' (Eigene Darstellung)

Nach Abschluss dieses Verarbeitungsschritts ist es möglich, die Daten auf der Analyseplattform einzusehen. Abbildung 46 zeigt die initiale Sicht auf die fünfzehn vorfallsrelevanten docx-Dateien. Die Indexierung der Ansicht ist in diesem Fall an dem Erstellungsdatum der einzelnen Dateien orientiert. Eine (Neu-)Indexierung bietet indes

die Möglichkeit, den Index an dem letzten Dateizugriff oder der letzten Dateiänderung auszurichten (Farmer 2000).

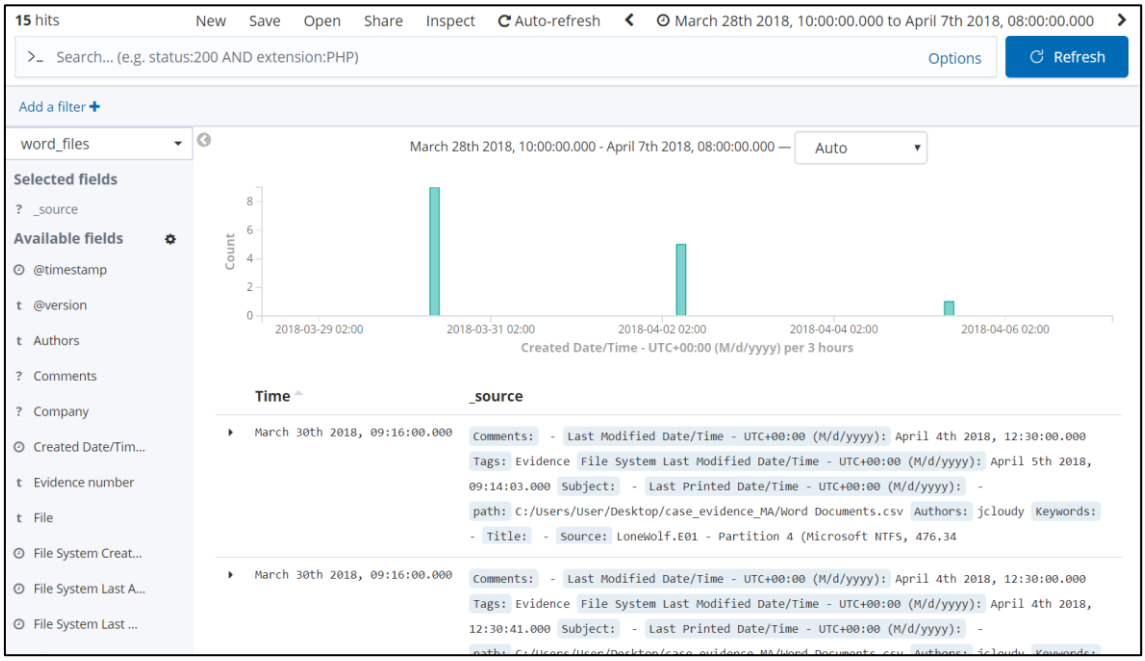

Abbildung 46: Darstellung der docx-Dateien im Index ‚Word Documents' (Eigene Darstellung)

Ab diesem Zeitpunkt können der Inhalt, die Sortierung, der Zeitraum und die Darstellungsart beliebig angepasst werden, während die zugrunde liegende Datenstruktur und deren Inhalt unverändert in der Elasticsearch-Instanz verbleiben. Diese Schritte bilden somit die Grundlage für die Erstellung von Darstellungen, die durch den Analysten oder Ermittler frei gestaltet und dadurch interaktiv angepasst werden können. Um alle vorfallsrelevanten Daten, die während der Untersuchung exportiert wurden, darzustellen, wurden alle der vierzehn Datensammlungen wie oben beschrieben erfasst, verarbeitet und gespeichert. Im Anschluss an diese Tätigkeit ist es möglich, Daten und Informationen aus einem oder mehreren der hierzu erstellten, im Folgenden angeführten Indizes grafisch für die Analysetätigkeit aufzubereiten:

- all_desktop_files
- aws_keys
- chrome_autofill
- chrome_current_session
- chrome_current_tabs

- chrome_logins
- classifieds_urls
- file_system_items
- google_searches
- parsed_search_queries
- pdf_documents
- pictures
- powerpoint_files
- word_files

In der Analysephase ist es nicht möglich, einzelne Indizes und deren Inhalte nachträglich zu verändern, Elemente hinzuzufügen oder zu entfernen. Die Anzahl an Indizes, die innerhalb dieser Architektur erstellt werden können, ist grundsätzlich nicht beschränkt. Maßgebend sind die verfügbare Speicherkapazität und die Leistungsfähigkeit der Elasticsearch-Instanz bzw. des Clusters.

8.3 Visualisierung der Vorfallsdaten

Um zu bestimmen, ob die Aktionen, die zur Planung dieser fiktiven Straftat ausgeführt wurden, das Werk eines Einzeltäters oder einer Gruppe von Tätern sind, kann zunächst der Zeitpunkt bestimmt werden, an dem die Planung begonnen wurde. Zusätzlich ist es möglich, zu Beginn der Analyse den geplanten oder tatsächlichen Tatzeitpunkt zu bestimmen. Der Zeitraum zwischen diesen beiden Punkten ermöglicht es, ausgehend von der Planungstätigkeit oder von dem geplanten bzw. tatsächlichen Tatzeitpunkt einzelne Akteure zu identifizieren.

Abbildung 47 zeigt auf Grundlage der vorfallsrelevanten Daten eine zeitliche Übersicht über die Planungstätigkeit.

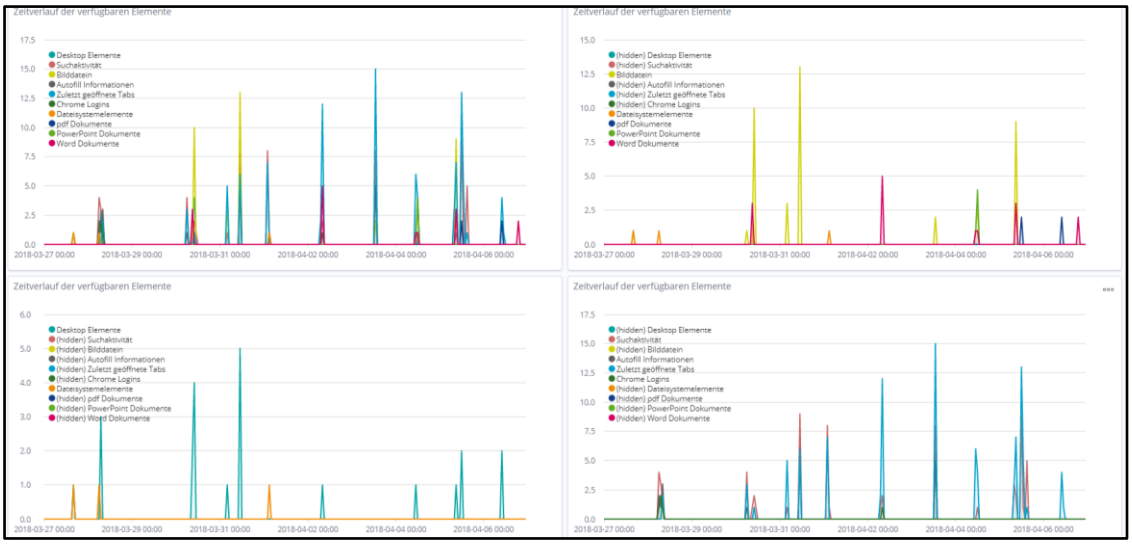

Abbildung 47: Zeitliche Übersicht der Planungstätigkeit (Eigene Darstellung)

Der Startzeitpunkt, 27. März 2018, ergibt sich aus dem Datum, an dem das Quellsystem in Betrieb genommen wurde. Das Ende der Planungstätigkeit, der 6. April 2018, ist der Zeitpunkt der Festnahme des Anwenders. Der Zeitverlauf umfasst zunächst nur die Planungstätigkeit und lässt ggf. Rückschlüsse auf das Nutzungsverhalten des Anwenders während der Planung zu. Um diese zu verifizieren, müssen die Daten durch beschreibende Informationen ergänzt werden. In Abbildung 48 sind die Zeitverläufe, durch Kontextinformationen angereichert, dargestellt.

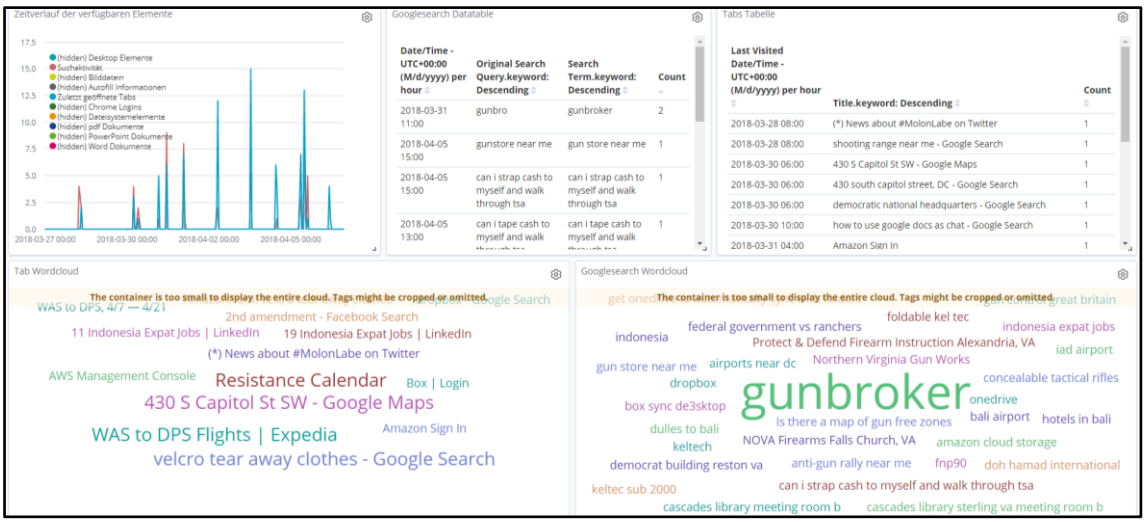

Abbildung 48: Darstellung des Zeitverlaufs, angereichert durch Kontextinformationen (Eigene Darstellung)

Ausgehend von dieser Darstellung kann die Datenanalyse auf einen Zeitraum oder ein Element aus einer Stich- und Suchwortliste eingegrenzt werden. Abbildung 49 stellt den Zeitraum zwischen 00:00 und 16:00 Uhr am 28. März 2018 dar.

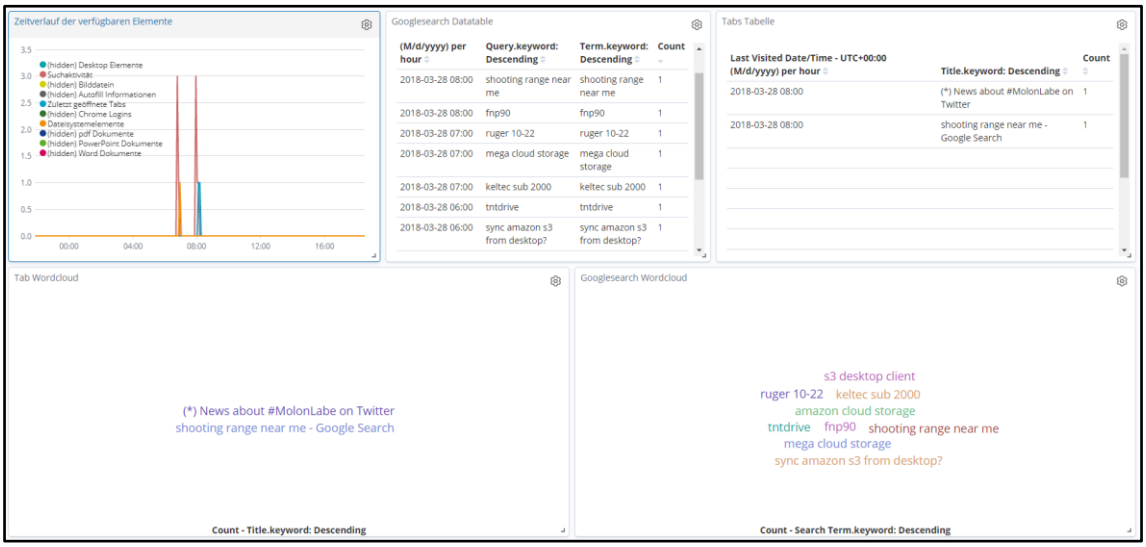

Abbildung 49: Darstellung des Ereigniszeitraums am 28 März 2018 zwischen 00 und 16 Uhr (Eigene Darstellung)

Diese Informationen sind erste Indikatoren zur Beschreibung des Vorfalls und der damit verbundenen Planungstätigkeit. Vor dem Hintergrund, dass es sich bei diesem fiktiven Vorfall um eine geplante Massenerschießung handelt, könnten die hier dargestellten Informationen bereits als Grundlage für eine weiterführende Analysetätigkeit dienen. Die Darstellung und Integration weiterer Datenquellen wird in Abbildung 50 aufgezeigt. Die darin enthaltenen Informationen können zunächst weiter aufbereitet oder unter Berücksichtigung relevanter Dateitypen und Bezeichnungen sortiert und gefiltert werden.

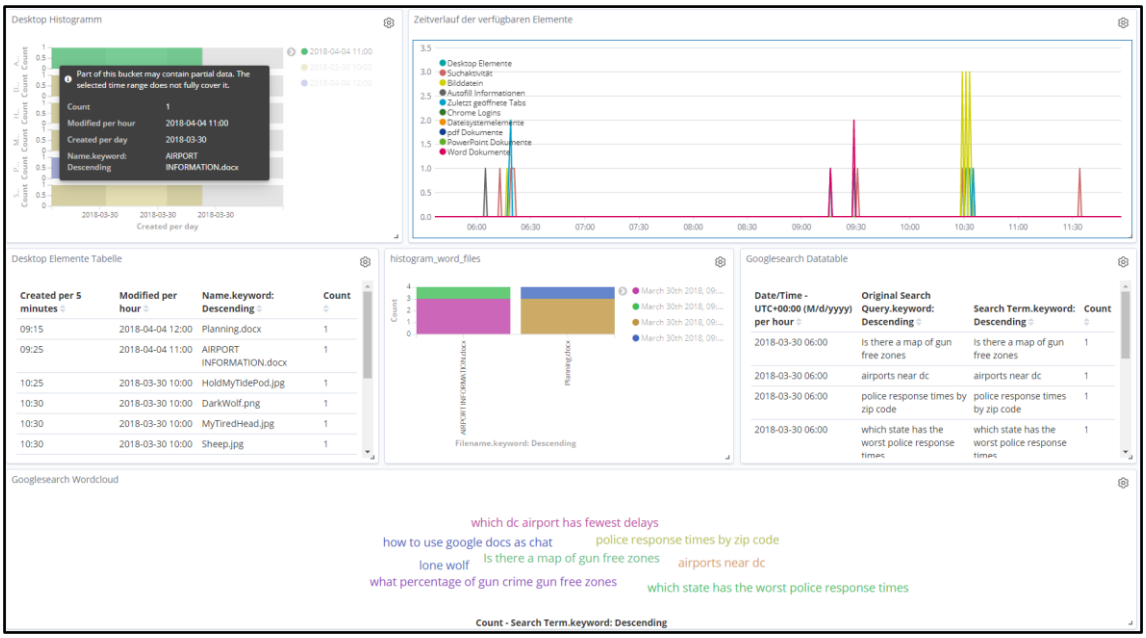

Abbildung 50: Korrelation von Dateitypen und -bezeichnungen (Eigene Darstellung)

Nun ist es möglich, relevante Dateitypen zu selektieren und diese, z. B. nach deren Häufigkeit, Inhalt oder Ablageort, weiter zu untersuchen. In Abbildung 51 werden die verfügbaren Elemente innerhalb des untersuchten Dateisystems dargestellt.

Abbildung 51: Darstellung der Dateisystemelemente (Eigene Darstellung)

An dieser Abbildung ist erkennbar, dass vier .docx-Dateien innerhalb des Dateisystems existieren, die in bis zu fünf verschiedenen Dateisystemverzeichnissen abgelegt wurden.

Diese Information wird durch das Datum der Erstellung und den Zeitpunkt der letzten Änderung ergänzt. Durch die Darstellung dieser Metadaten ist es möglich, den Autor und die Historie (Erstellung und Änderungen) der Dokumente nachzuvollziehen. Abbildung 52 beinhaltet ein Dendrogramm, an dessen Wurzelknoten der Autor der Dokumente zu finden ist.

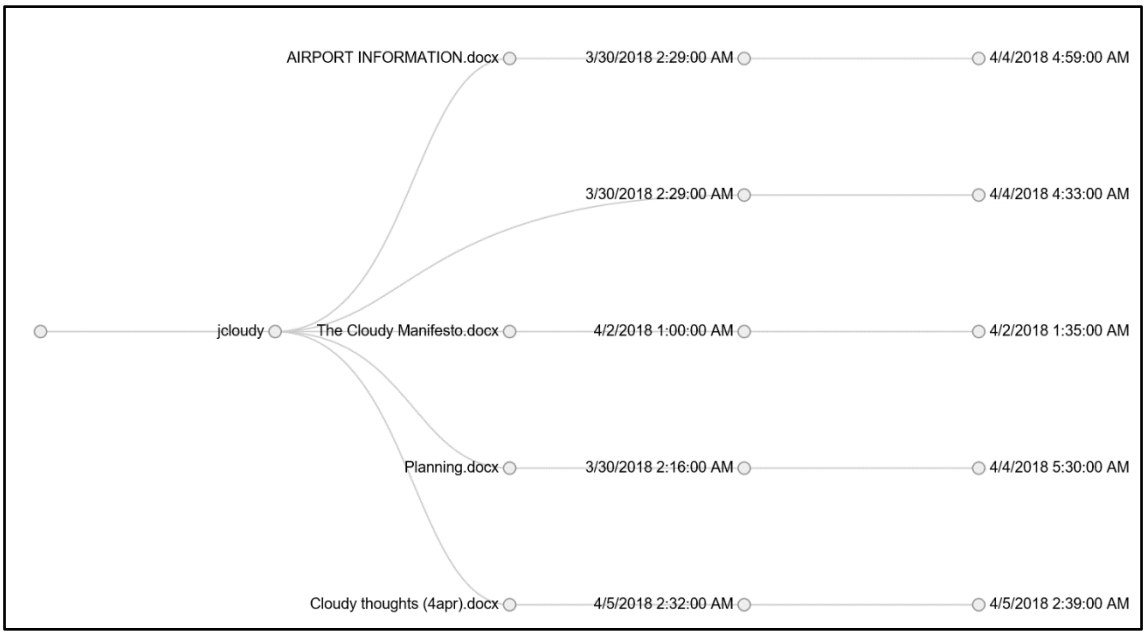

Abbildung 52: Autorenbestimmung der .docx-Dateien (Eigene Darstellung)

Informationen über die, durch den Anwender abgesetzten, WWW-Suchen werden, wie in Abbildung 53 dargestellt, für eine detaillierte Analyse durch den vollständigen Suchbegriff bzw. die Suchphrase, die Mengenverteilung und den Zeitverlauf ergänzt.

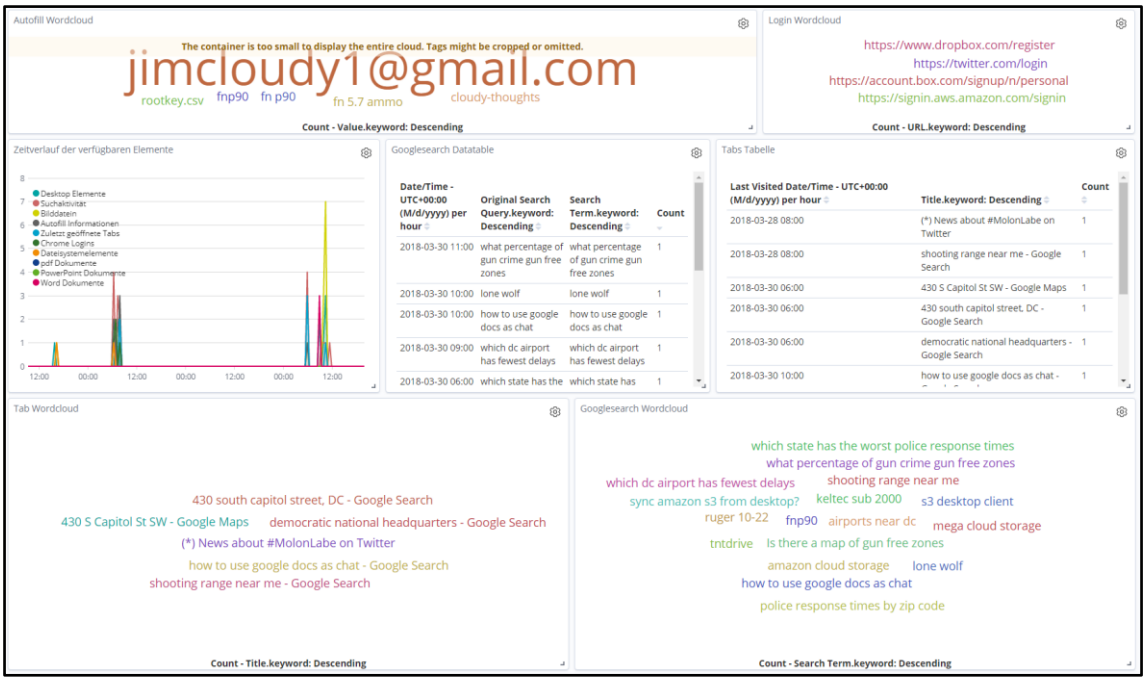

Abbildung 53: Analyse der Nutzungshistorie durch Suchbegriffe (Eigene Darstellung)

Die entsprechenden Informationen ermöglichen es, das Nutzungsverhalten des Anwenders während der Planung nachzuvollziehen. Die Elemente der Wortwolke bieten in Verbindung mit dem verfügbaren Zeitverlauf die Möglichkeit, Zielsetzungen oder Fragestellungen des Anwenders zu bestimmen. Durch die Verknüpfung von autofill-Informationen, die innerhalb des verwendeten Browsers abgelegt werden und aus dem Nutzungsverhalten des Anwenders resultieren, und verfügbaren Dateisysteminformationen ist es möglich, Dokumente und Informationen zu identifizieren, die aufgrund der vermehrten Nutzung für den Anwender von Bedeutung sein könnten. Die Orientierung an diesen Artefakten, die innerhalb des Dateisystems abgelegt wurden, bildet die Grundlage für eine fallorientierte Stich- und Suchwortliste.

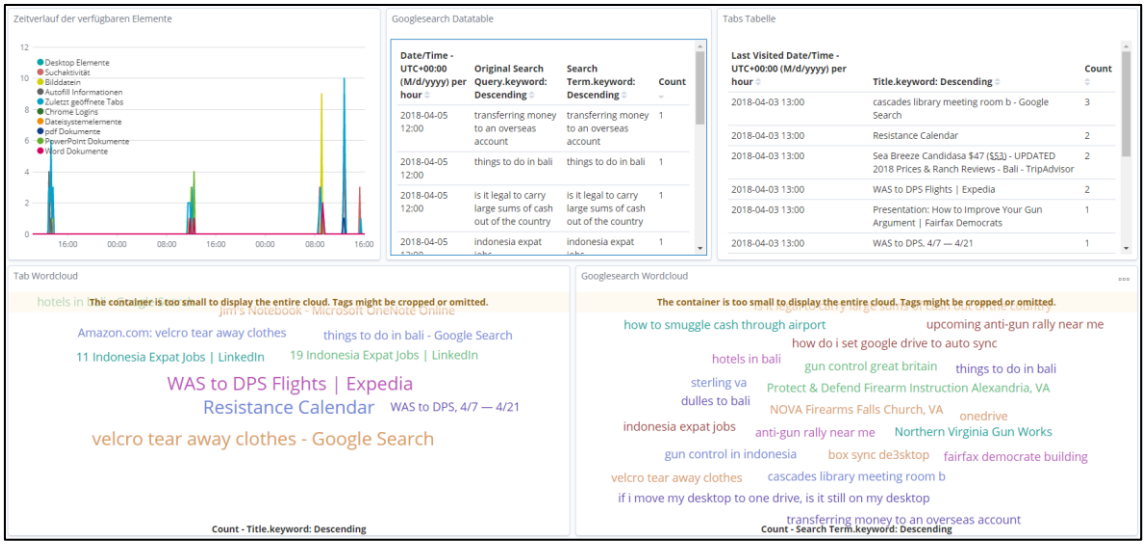

Abbildung 54: Darstellung von Elementen der Nutzungshistorie (Eigene Darstellung)

In Bezug auf diesen fiktiven Vorfall können auf Basis der interaktiven Grafik in Abbildung 53 und 54 die Küstenstadt Candidasa auf der indonesischen Insel Bali sowie verschiedene Waffen- und Munitionsarten als Bestandteil der Stich- und Suchwortliste identifiziert werden. Aus diesen Informationen, die als Elemente der Stich- und Suchwortliste bestimmt wurden, lässt sich der Planungsverlauf und das gewählte Vorgehen des Beschuldigten ableiten. Hierzu wurden aus dem Datenträgerabbild die folgenden Dokumente extrahiert:

- Planning.docx
- AIRPORT INFORMATION.docx
- The Cloudy Manifesto.docx
- Cloudy thoughts (4apr).docx

Die chronologische Durchsicht der Dokumente nach dem Erstellungs- und Änderungsdatum sowie die Indexierung der entsprechende Inhalte ermöglichen die Bestimmung weiterer fallrelevanter Informationen. Die folgende Abbildung 55 zeigt dieses Vorgehen anhand des Dokuments ‚Planning.docx ‘.

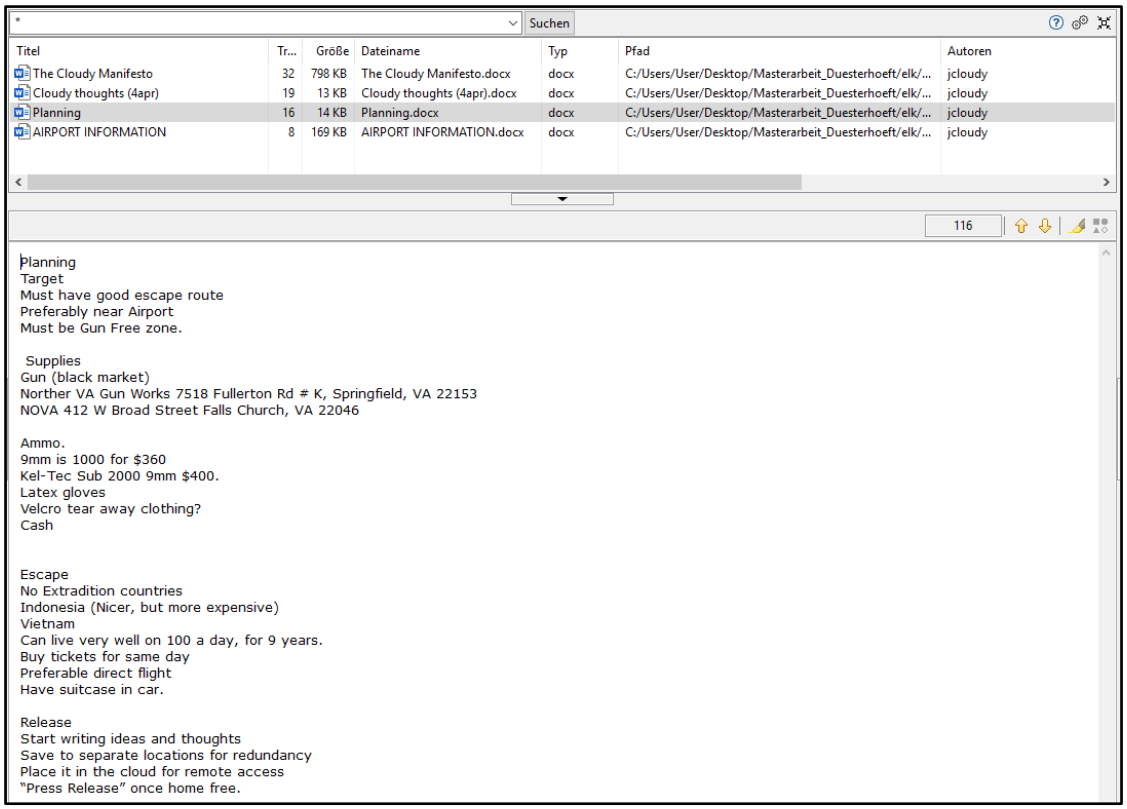

Abbildung 55: Indexierung des Dokuments ‚Planning.docx' und Volltextsuche (Eigene Darstellung)

Analog hierzu zeigt Abbildung 56 das Vorgehen bzgl. des Dokuments ‚Cloudy thoughts (4apr).docx'.

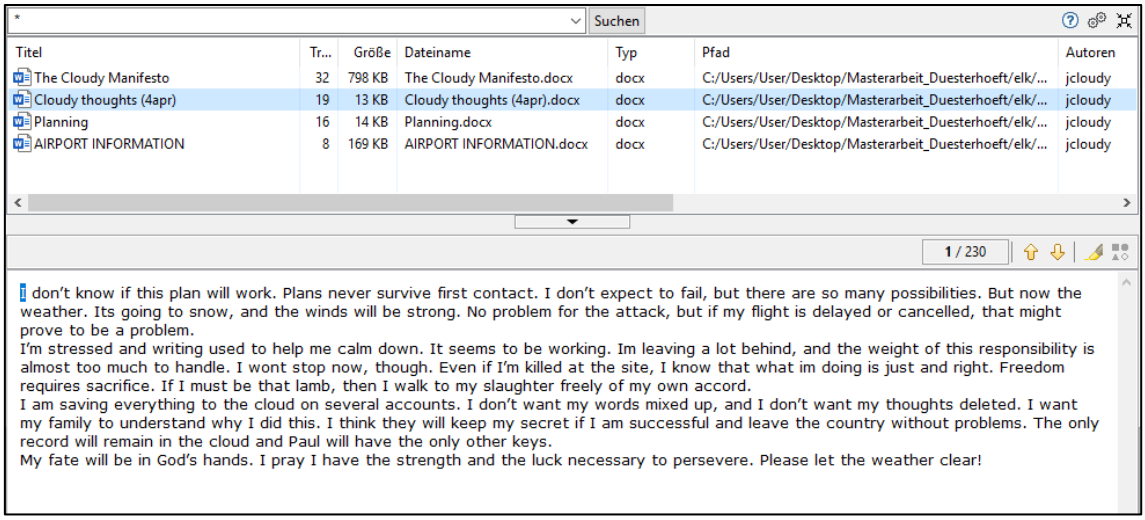

Abbildung 56: Indexierung des Dokuments ‚Cloudy thoughts (4apr).docx' und Volltextsuche (Eigene Darstellung)

Aus dieser Analysephase lassen sich weitere Elemente der Sich- und Suchwortliste ableiten. Wird diese Liste nun auf kommunikationsrelevante Elemente eingeschränkt, in diesem Fall auf die Begriffe ‚drive‘, ‚chat‘ und ‚cloud‘ ergibt sich das in Abbildung 57 dargestellte Resultat.

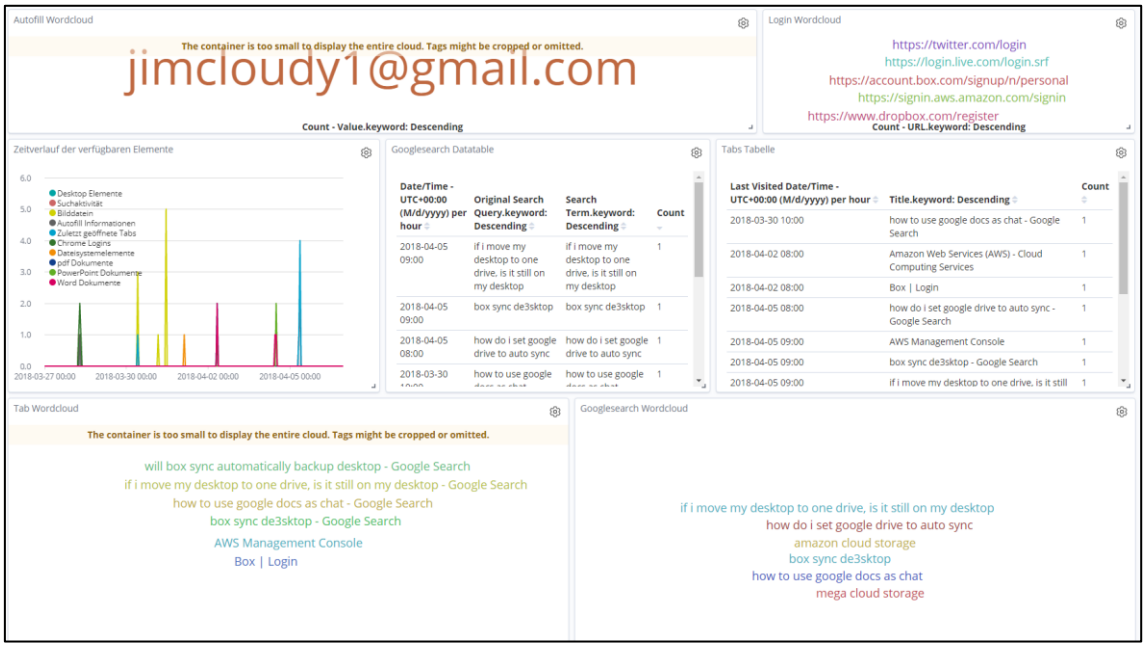

Abbildung 57: Suche nach Elementen der Stich- und Suchwortliste (Eigene Darstellung)

Vor dem Hintergrund der Suchanfrage ‚how to use google docs as chat‘, die am 30. März 2018 übermittelt wurde, kann eine .ldb-Datei, die mit dem Dokument ‚Brother Chat.gdoc‘ verknüpft ist, extrahiert werden. Ein Ausschnitt dieser Datei ist in Abbildung 58 dargestellt.

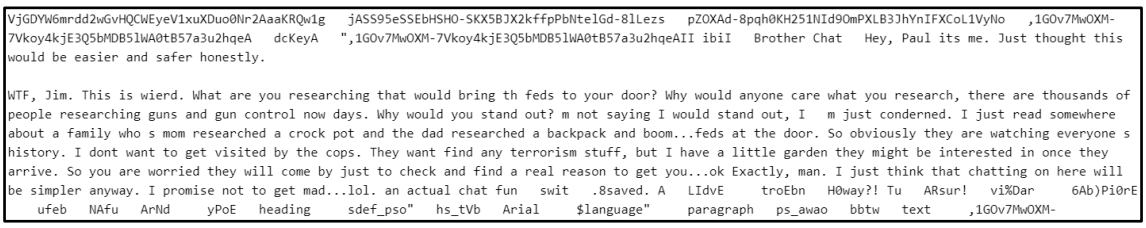

Abbildung 58: Inhalt der Datei ‚Brother Chat.gdoc‘ (Eigene Darstellung)

Auf Grundlage der Daten, die sich auf dem Endgerät befinden, kann die Frage danach, ob es sich um einen Einzeltäter oder eine Gruppe von Tätern handelt, eindeutig beantwortet werden: Die Hintergründe dieses Vorfalls sowie die Planungstätigkeit lassen sich auf die fiktive Person ‚Jim Cloudy‘ zurückführen. Begründet wird diese Aussage

durch die fehlende Kommunikation mit weiteren beteiligten Personen. Die auf dem Gerät verfügbare Kommunikation hinsichtlich des Vorfalls und dessen Planung beschränkt sich auf jene mit dem Hinweisgeber selbst, dem Bruder des Beschuldigten ‚Paul Cloudy'. Auch die Frage danach, was wo zu welchem Zeitpunkt wie passiert ist, sowie die Frage nach dem Ursprung des Vorfalls, also wodurch bzw. durch wen wurde der Vorfall verursacht, konnten unter Verwendung von interaktiven Grafiken abschließend beantwortet werden.

9 Einschränkungen und Herausforderungen bei der Daten- und Informationsvisualisierung

Die Herausforderungen bei der Daten- und Informationsvisualisierung während der forensischen Datenanalyse ergeben sich zunächst aus den Fragen nach den möglichen Datenquellen und der Datenstruktur, in der die Daten in den Systemen abgelegt wurden. Von Bedeutung für die Umsetzung der Datenanalyse unter Verwendung von interaktiven Grafiken sind jedoch auch Fragestellungen, die auf die Effektivität der Auswertungs- und Analysemöglichkeiten abzielen. Vor dem Hintergrund steigender Datenmengen, der Digitalisierung von nicht-elektronischen Artefakten, der Normalisierung von Datenquellen und der hohen Anforderungen bzgl. der Datenbereitstellung gilt es, Vorgehensmodelle zu erproben, mit denen diese Fragestellungen zufriedenstellend beantwortet und die auf Basis einer Vorfallssammlung zielgerichtet weiterentwickelt werden können. Interaktive Grafiken können innerhalb der Analysephase dazu verwendet werden, zentrale Fragestellungen einer forensischen Untersuchung zu beantworten. Sie zeigen vorfallsrelevante Informationen, deren Abhängigkeiten sowie Ursprung allgemein verständlich und können deshalb zur Daten- und Informationsanalyse eingesetzt werden.

Diese Form der Analyse ist jedoch nur dann vollständig möglich, wenn die vorhandene Datenstruktur die Verknüpfung von kontextbasierten Daten mit relevanten Ereignissen und Informationen ermöglicht und unterstützt. Hierbei gilt, dass nur das dargestellt werden kann, was auch da ist. Eine inhaltliche Verknüpfung von zwei oder mehr Datenquellen ist also prinzipiell auf Grundlage eines beliebigen Attributs möglich. Erfolgt die Verknüpfung zweier Datenquellen über ein inhaltliches Attribut, z. B. ein Element aus einer Stich- oder Suchwortliste, besteht die Möglichkeit, dass vorfallsrelevante Informationen unberücksichtigt bleiben, da diese durch das gewählte Attribut nicht erfasst werden. Um auszuschließen, dass einzelne Elemente aufgrund bestehender Verknüpfungen zwischen verschiedenen Datenquellen während der Datenanalyse nicht berücksichtigt werden, ist es möglich, die Ereigniszeit als Verbindungselement einzusetzen. Damit die Ereigniszeit diesen Zweck erfüllen kann, ist es notwendig, dass die Zeitinformation über alle Datenquellen hinweg normalisiert werden können.

Unabhängig davon, ob es sich bei den Daten aus den Quellsystemen um strukturierte, semi- oder unstrukturierte Daten handelt, kann die Notwendigkeit der Datentransformation bestehen. Diese Maßnahme beinhaltet z. B. die Attribution von Elementen, Ereignissen oder Objekten, die durch das Quellsystem erstellt, verwaltet oder verarbeitet werden. Diese Maßnahme ist vor allem von Bedeutung, da innerhalb des forensischen Prozesses grundsätzlich jedes verfügbare Artefakt potenziell zur Aufklärung des Vorfalls beiträgt. Es muss daher möglich sein, jede Datei oder Information auf eine Analyseplattform überführen zu können, ohne sie dabei inhaltlich zu verändern. Ziel ist es, die Information, die sich aus einer Datei oder beliebigen Daten ergibt, mit anderen Daten zu verknüpfen. Sofern es nicht möglich ist, alle einen Vorfall betreffenden Daten und Informationen vollständig auf eine Analyseplattform zu übertragen, ist es auch nicht möglich, die Datenanalyse unter Verwendung von interaktiven Grafiken, die mittels dieser Plattform erstellt werden, vollständig durchzuführen und abzuschließen. Daraus ergibt sich die Notwendigkeit, für die forensische Datenanalyse auf weitere Maßnahmen, Anwendungen und Werkzeuge zurückzugreifen, um zusätzliche Erkenntnisse zur Aufklärung eines Vorfalls zu generieren.

10 Fazit

Das Ziel dieser Arbeit war es, Methoden und Schnittstellen zu bestimmen, die es ermöglichen, forensische Datenanalysen mithilfe interaktiver Grafiken durchzuführen. Diese können insbesondere in der Untersuchungs- und Analysephase durch Ermittler und Analysten eingesetzt werden, um Untersuchungsergebnisse und Nachweise zu ermitteln und zielgruppenorientiert dazustellen.

In Kapitel 10.1 werden die Arbeit und die dafür durchgeführten Aktivitäten und Entwicklungen zusammengefasst. Daran schließt in Kapitel 10.2 eine Bewertung des dargestellten Vorgehens zur forensischen Datenanalyse unter Verwendung von interaktiven Grafiken zur Daten- und Informationsvisualisierung an. Abgeschlossen wird dieses Fazit mit Kapitel 10.3 durch einen Ausblick auf die Nutzungsmöglichkeiten von interaktiven Grafiken während der forensischen Datenanalyse.

10.1 Zusammenfassung

Zur Beantwortung der Forschungsfrage wurden bestehende Methoden und Werkzeuge zur forensischen Datenanalyse und Daten- und Informationsvisualisierung erfasst und in Bezug auf deren Einsatzgebiete beschrieben. Über die Bestimmung von Anforderungen, die an eine forensische Datenanalyse in Verbindung mit der Daten- und Informationsvisualierung gestellt werden, war es möglich, Visualisierungsmethoden zu bestimmen, die im Kontext der forensischen Datenanalyse eingesetzt werden können. In Verbindung mit der Fall- und Ereignisklassifikation wurden zwei Domänen mit jeweils fünf Vorfallsklassen bestimmt. Diese ermöglichen es, auftretende Vorfälle zu klassifizieren. Ausgehend von dieser Klassifikation ist es möglich, vorfallsrelevante Daten und Informationen während der forensischen Datenanalyse unter Verwendung von interaktiven Grafiken zu analysieren. Dieses Vorgehen zur Durchführung und Unterstützung der Analysetätigkeit war ein Bestandteil dieser Arbeit und konnte erfolgreich verprobt werden. Die Messbarkeit des zur Implementierung des Testfalls verwendeten Vorgehens und dessen Nutzen für die Datenanalyse ist nicht gegeben. Dennoch lässt sich aus den beschriebenen Erkenntnissen die Hypothese ableiten, dass die zur Analyse des Testvorfalls bereitgestellten interaktiven Grafiken Ermittler und

Analysten während der forensischen Datenanalyse aufgrund ihrer Struktur und Einsatzmöglichkeiten nachhaltig unterstützen können.

10.2 Bewertung

Das in dieser Arbeit beschriebene Vorgehen zur Nutzung von interaktiven Grafiken innerhalb einer forensischen Untersuchung kann in mehreren Prozessphasen eingesetzt werden. Die Erfassung von vorfallsrelevanten Ereignissen, Daten und Informationen über die Analyseplattform bildet die Voraussetzung für die Unterstützung der Prozessphasen der Untersuchung, Datenanalyse und Dokumentation. Die Erstellung der Fall- und Ereignisklassifikation ermöglicht die Klassifikation von Vorfällen und eine Handlungsempfehlung bzgl. der Darstellungsoptionen. Das Ziel ist es, die zentralen W-Fragestellungen (was, wo, wie, wann und wer) während der Analysephase zu beantworten. Der Reifegrad dieses Verfahrens ist insofern begrenzt, als dass sich die Nutzung auf einen einzigen vollständig implementierten Anwendungsfall beschränkt. Daher ist es wahrscheinlich, dass sich Informationen bzgl. möglicher Dateinamenserweiterungen oder des voraussichtlichen Datenvolumens im Nutzungsverlauf ändert. Mögliche Änderungen können aus Erfahrungswerten, die sich aus der Nutzung des Verfahrens ergeben, resultieren. Bereits bekannte Einschränkungen der Einsatzmöglichkeiten von interaktiven Grafiken sollten im Rahmen der Anwendungsanalyse, der Untersuchung von Computerarchitekturen oder dem Reverse Engineering berücksichtigt werden. Diese ergeben sich vor allem aus der hohen Komplexität der fachlichen Analysetätigkeit. Erkenntnisse, die aus einer solchen Analyse resultieren, können jedoch durch interaktive und Informationsgrafiken, z. B. zur Erkenntnisdokumentation, dargestellt werden. Die Erfahrungen, die sich aus der Nutzung des Verfahrens ergeben, bilden gleichzeitig die Grundlage für eine kontinuierliche Weiterentwicklung des Verfahrens. Daher ist es möglich, dass die existierende Fall- und Ereignisklassifikation zukünftig angepasst werden muss und sich daraus Änderungen an den verfügbaren Domänen oder vorhandenen Vorfallsklassen ergeben. Dennoch bietet der Einsatz von interaktiven Grafiken zur forensischen Datenanalyse die Möglichkeit, Daten und Informationen ganzheitlich darzustellen. Einzelne vorfallsrelevante Elemente können auf diese Weise allgemein verständlich dargestellt und im Kontext mit anderen

Informationen auch durch fachfremde Analysten und Ermittler bewertet werden. Diese Einsatzmöglichkeiten gilt es weiterhin praktisch zu erproben und zu verifizieren.

10.3 Ausblick

Das in dieser Arbeit vorgestellte Vorgehen zur forensischen Datenanalyse stellt eine Analysemethode dar, die im Rahmen einer forensischen Untersuchung durch das Auftreten beliebiger Vorfallsarten erweitert werden kann. Um die Aufwände für die Datenintegration und die Dokumentation der ausgeführten Dateioperationen zur Normalisierung und Transformation zu reduzieren, kann auf Grundlage der Fall- und Ereignisklassifikation eine normierte Datenstruktur je Element der Klassifikation bestimmt werden. Daten und Informationen, die im Rahmen der forensischen Datenanalyse zu untersuchen sind, werden dann zentral über einen Extract, Transform, Load-Prozess (ETL) in diese Struktur überführt. Dieser Prozess beinhaltet die Integration und Validierung der erfassten Daten. Eine Bereinigung oder Transformation sollte dabei ebenso wie die Fehlerbehandlung in Abstimmung mit dem Analysten oder Ermittler erfolgen. Dies gilt insbesondere vor dem Hintergrund der Datenverwendung. Sowohl fallspezifische Aufwände zur Vorbereitung der Datenanalyse als auch die Wahrscheinlichkeit, dass einzelne Arbeitsschritte unvollständig dokumentiert und damit nur bedingt wiederholbar sind, lassen sich durch diese Struktur vermeiden.

Zur Umsetzung dieser vorbereitenden Maßnahme werden Schnittstellen zu vor- und nachgelagerten Teilprozessen innerhalb des forensischen Prozesses benötigt. Diese Schnittstellen dienen dem Austausch von Erkenntnissen, die während der Analyse ermittelt und dokumentiert werden, der Verarbeitung von Daten aus verschiedenen Datenquellen über die Prozessphasen der Datensammlung und Untersuchung hinweg sowie erfüllen technische und organisatorische Anforderungen, die sich aus landes- oder organisationsspezifischen Anforderungen an die forensische Untersuchung ergeben. Eine prototypische Implementierung dieser Analysemethode besteht aus einer definierten Fall- und Ereignisklassifikation, die es ermöglicht, Vorfallsdaten zeitnah zur Verfügung zu stellen, und einer passenden Analyseplattform, die sowohl ortsunabhängig als auch losgelöst von der Fallgröße eine systemseitige Grundlage für die Datenanalyse bietet. Beide Komponenten können sowohl gemeinsam als auch isoliert weiterentwickelt

werden. Für die Bereitstellung der Analyseplattform zur Datenanalyse während einer forensischen Untersuchung sollte ein ‚Golden Image' erstellt werden. Dieses Image ermöglicht die Analyse von Daten, die während des Vorfalls angefallen sind, in einer dedizierten, von anderen Vorfällen getrennten Umgebung und verkürzt die Bereitstellungszeit der Analyseplattform. Die Protokollierung der Analysetätigkeit und anderen Sitzungsaktivitäten kann durch eine Sitzungsaufzeichnung erfolgen. Um Ermittler und Analysten in der Analysetätigkeit zu unterstützen, kann eine Sammlung von interaktiven Grafiken und Vorlagen zur Daten- und Informationsvisualisierung im Golden Image hinterlegt werden. Beispiele für derartige Grafiken und Vorlagen können folgende sein:

- geografische Übersichtskarten, die eine Nutzung und Verbreitung von vorfallsrelevanten Daten beschreiben,
- eine Ereignisübersicht, die beschreibt, zu welchem Zeitpunkt und durch welchen Prozess oder Anwender ein Ereignis ausgelöst wurde,
- Verbindungsdiagramme, die eine Beziehung zwischen zwei oder mehreren Attributen aufzeigen sowie
- eine Darstellung der Lösch- und Änderungshistorie, um eine Übersicht über Änderungen an Dateisystemelementen zu geben.

Die Nutzung von interaktiven Grafiken in der Analysephase einer forensischen Untersuchung zeichnet sich nach der Integration der hier beschriebenen Maßnahmen vor allem durch eine hohe Funktionalität und die Bedien- und Anpassbarkeit aus. Wird zur Implementierung der Analyseplattform und zum Betrieb der Analyseinfrastruktur ausschließlich quelloffene Software eingesetzt, ist es möglich, eine allgemein zugängliche Struktur zu erstellen und zu betreiben (Garfinkel 2013). Dieses Vorgehen ermöglicht eine lösungsorientierte Analysetätigkeit, die durch eine standardkonforme Dokumentation der sich daraus ergebenden Resultate den Anforderungen an die forensische Datenanalyse entspricht.

Abkürzungsverzeichnis

BSI	Bundesamt für Sicherheit in der Informationstechnik
CERT	Computer-Emergency-Response-Team
eDiscovery	Electronic discovery
EDRM	Electronic-Discovery-Reference-Model
ETL	Extract, Transform, Load Prozess
FBI	Federal Bureau of Investigation
FTK	Forensic Toolkit
HTTP	Hypertext-Transfer-Protocol
HTTPS	Hypertext-Transfer-Protocol-Secure
IETF	Internet Engineering Task Force
IOCE	International Organisation on Computer Evidence
PST	Personal Store
RFC	Request for Comments
SIEM	Security-Information-and-Event-Management-System
SNMP	Simple Network Management Protocol
WWW	World Wide Web

Abbildungsverzeichnis

Abbildung 1: Austauschprinzip nach Dr. Edmond Locard (Eigene Darstellung, in Anlehnung an: (Locard 1930)) 2

Abbildung 2: Der forensische Prozess nach BSI (Eigene Darstellung, in Anlehnung an: (BSI 2011, S. 84)) 11

Abbildung 3: CERT-Taxonomie (Eigene Darstellung, in Anlehnung an: (BSI 2011, S. 29)) 11

Abbildung 4: Vorgehen zur Vorfalluntersuchung (Eigene Darstellung) 16

Abbildung 5: Zeitstrahl zur Darstellung eines strafrechtlichen Verstoßes (Eigene Darstellung) 20

Abbildung 6: Balkendiagramm zu den Importen und Exporten Schottlands aus dem Jahr 1781 (Playfair 1786) 27

Abbildung 7: Bildstatistik zur Gesamtzahl der Güterverkehrsleistungen mittels Eisenbahnen aus dem Jahr 1914 (Brinton 1914, S. 40) 27

Abbildung 8: Informationsvisualisierung am Beispiel bestätigter Cholerainfektionen (Barton 2018) 31

Abbildung 9: Logmanagement Dashboard (Eigene Darstellung) 36

Abbildung 10: Phasen des Electronic-Discovery-Reference-Model (Eigene Darstellung, in Anlehnung an: (EDRM Duke Law 2019)) 38

Abbildung 11: eDiscovery – alle Elemente einer Personal-Store-Datei (Eigene Darstellung) 40

Abbildung 12: : eDiscovery – ausgewählte Elemente einer Personal-Store-Datei (Eigene Darstellung) 40

Abbildung 13: Visualisierung von HTTP- und HTTPS-Aktivitäten mit Hviz (Gugelmann et al. 2015) 46

Abbildung 14: Überprüfung von ‚www.bbc.co.uk' mit Lookyloo (Eigene Darstellung, in Anlehnung an: (CIRCL.lu 2019)) 47

Abbildung 15: Ereignisvisualisierung unter Berücksichtigung des Eintrittszeitpunkts und Orts mit GeoTime (GeoTime 2019) 48

Abbildung 16: Tabellarische Darstellung zur Analyse des Dateipfades (Eigene Darstellung) 49

Abbildung 17: Tabellarische Darstellung zur Analyse des Dateipfades, erweitert durch die Attribute MD5 und SHA1 (Eigene Darstellung) 49

Abbildung 18: Zeitachse zur Ablage von inkriminierten Inhalten auf einem Dateiserver (Eigene Darstellung) 50

Abbildung 19: Datenhierarchie zur Analyse von Wortgruppen (Eigene Darstellung) 51

Abbildung 20: Informationsgrafik zur Vorfallsdarstellung (Eigene Darstellung) 56

Abbildung 21: Visualisierungsmethode – Verlaufsdiagramm (Eigene Darstellung) 75

Abbildung 22: Visualisierungsmethode – gruppiertes Balkendiagramm (Eigene Darstellung) 75

Abbildung 23: Visualisierungsmethode – Blasendiagramm (Eigene Darstellung) 76

Abbildung 24: Visualisierungsmethode – Matrix-Diagramm (Eigene Darstellung) 76

Abbildung 25: Visualisierungsmethode – Baumdiagramm (Eigene Darstellung) 77

Abbildung 26: Visualisierungsbestimmung – Netzwerkdiagramm (Eigene Darstellung) 78

Abbildung 27: Visualisierungsbestimmung – Sehnendiagramm (Eigene Darstellung) 78

Abbildung 28: Visualisierungsmethode – Punktverteilungskarte (Eigene Darstellung) 79

Abbildung 29: Visualisierungsmethode – Verbindungskarte (Eigene Darstellung) 80

Abbildung 30: Visualisierungsmethode – Blasenkarte (Eigene Darstellung) 80

Abbildung 31: Visualisierungsmethode – Wortwolke (Eigene Darstellung) 81

Abbildung 32: Visualisierungsmethode – Dendrogramm (Eigene Darstellung) 81

Abbildung 33: Visualisierungsmethode – Zeitleiste (Eigene Darstellung) 82

Abbildung 34: Visualisierungsmethode - Gantt-Diagramm (Eigene Darstellung) 82

Abbildung 35: Visualisierungsauswahl für die Fallklasse ‚Erpressung' aus der Domäne Computerkriminalität (Eigene Darstellung) 85

Abbildung 36: Visualisierungsauswahl für die Fallklasse ‚Spionage' aus der Domäne Computerkriminalität (Eigene Darstellung) 86

Abbildung 37: Visualisierungsauswahl für die Fallklasse ‚Phishing' aus der Domäne Computerkriminalität (Eigene Darstellung) 87

Abbildung 38: Visualisierungsauswahl für die Fallklasse ‚Sabotage' aus der Domäne Computerkriminalität (Eigene Darstellung) 88

Abbildung 39: Visualisierungsauswahl für die Fallklasse ‚Missbrauch' aus der Domäne Computerkriminalität (Eigene Darstellung) 89

Abbildung 40: Visualisierungsauswahl für die Fallklasse ‚Geldwäsche' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung) 90

Abbildung 41: Visualisierungsauswahl für die Fallklasse ‚Korruption' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung) 91

Abbildung 42: Visualisierungsauswahl für die Fallklasse ‚Insiderhandel' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung) 92

Abbildung 43: Visualisierungsauswahl für die Fallklasse ‚Insolvenzverfahren' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung) 93

Abbildung 44: Visualisierungsauswahl für die Fallklasse ‚Marktmanipulation' aus der Domäne Wirtschaftskriminalität (Eigene Darstellung) 94

Abbildung 45: Programmanweisung zur Erfassung, Verarbeitung und Normalisierung des Index ‚Word Documents' (Eigene Darstellung) 99

Abbildung 46: Darstellung der docx-Dateien im Index ‚Word Documents' (Eigene Darstellung) 100

Abbildung 47: Zeitliche Übersicht der Planungstätigkeit (Eigene Darstellung) 102

Abbildung 48: Darstellung des Zeitverlaufs, angereichert durch Kontextinformationen (Eigene Darstellung) 102

Abbildung 49: Darstellung des Ereigniszeitraums am 28 März 2018 zwischen 00 und 16 Uhr (Eigene Darstellung) 103

Abbildung 50: Korrelation von Dateitypen und -bezeichnungen (Eigene Darstellung) 104

Abbildung 51: Darstellung der Dateisystemelemente (Eigene Darstellung) 104

Abbildung 52: Autorenbestimmung der .docx-Dateien (Eigene Darstellung) 105

Abbildung 53: Analyse der Nutzungshistorie durch Suchbegriffe (Eigene Darstellung) 106

Abbildung 54: Darstellung von Elementen der Nutzungshistorie (Eigene Darstellung) 107

Abbildung 55: Indexierung des Dokuments ‚Planning.docx' und Volltextsuche (Eigene Darstellung) 108

Abbildung 56: Indexierung des Dokuments ‚Cloudy thoughts (4apr).docx' und Volltextsuche (Eigene Darstellung) 108

Abbildung 57: Suche nach Elementen der Stich- und Suchwortliste (Eigene Darstellung) 109

Abbildung 58: Inhalt der Datei ‚Brother Chat.gdoc' (Eigene Darstellung) 109

Tabellenverzeichnis

Tabelle 1: Informationsquellen im Rahmen der technischen Fehleranalyse 22

Tabelle 2: Fall- und Ereignisklassifikation – Computerkriminalität: Erpressung 64

Tabelle 3: Fall- und Ereignisklassifikation – Computerkriminalität: Spionage 65

Tabelle 4: Fall- und Ereignisklassifikation – Computerkriminalität: Phishing 66

Tabelle 5: Fall- und Ereignisklassifikation – Computerkriminalität: Sabotage 67

Tabelle 6: Fall- und Ereignisklassifikation – Computerkriminalität: Missbrauch 68

Tabelle 7: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Geldwäsche 69

Tabelle 8: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Korruption.....70

Tabelle 9: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Insiderhandel 71

Tabelle 10: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Insolvenzverfahren 72

Tabelle 11: Fall- und Ereignisklassifikation – Wirtschaftskriminalität: Marktmanipulation 73

Quellen- und Literaturverzeichnis

AccessData Forensic Toolkit (FTK) (2017): Forensic Toolkit Imager. Online verfügbar unter https://accessdata.com/product-download/ftk-imager-version-4.2.0, zuletzt geprüft am 03.04.2019.

Ahn, Jae-wook; Plaisant, Catherine; Shneiderman, Ben (2011): A Task Taxonomy of Network Evolution Analysis. Online verfügbar unter http://www.cs.umd.edu/hcil/trs/2011-09/2011-09.pdf, zuletzt geprüft am 20.03.2018.

Al-Zaidy, Rabeah; Fung, Benjamin C. M.; Youssef, Amr M. (2011): Towards discovering criminal communities from textual data. In: William Chu (Hg.): Proceedings of the 2011 ACM Symposium on Applied Computing. the 2011 ACM Symposium. TaiChung, Taiwan. ACM Special Interest Group on Applied Computing. New York, NY: ACM, S. 172.

Balakrishnan, Balaji; Walker, Chris (2014): Security Data Visualization. Online verfügbar unter https://www.sans.org/reading-room/whitepapers/metrics/security-data-visualization-36387, zuletzt geprüft am 05.04.2019.

Barton, Marc (2018): JOHN SNOW AND THE 1854 CHOLERA OUTBREAK. Online verfügbar unter https://www.pastmedicalhistory.co.uk/john-snow-and-the-1854-cholera-outbreak/, zuletzt aktualisiert am 25.03.2019, zuletzt geprüft am 25.03.2019.

Baryamureeba, Venansius; Tushabe, Florence (2004): The Enhanced Digital Investigation Process Model. Online verfügbar unter http://dfrws.org/sites/default/files/session-files/paper-the_enhanced_digital_investigation_process_model.pdf, zuletzt geprüft am 25.04.2019.

Beebe, Nicole (2009): Digital Forensic Research. The Good, the Bad and the Unaddressed. In: Gilbert Peterson und Sujeet Shenoi (Hg.): Advances in Digital Forensics V. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 17–36.

BKA (2003): nformationsund Kommunikationskriminalität. Eine Literaturauswahl. Bundeskriminalamt. Online verfügbar unter https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/Cod

Literaturreihe/8_15_Informations-undKommunikationskriminalitaet.pdf?__blob=publicationFile&v=2, zuletzt geprüft am 04.05.2019.

Brinton, Willard C. (1914): Graphic Methods For Presenting Facts. Online verfügbar unter https://aviz.fr/wiki/uploads/Bertifier/brinton-graphicMethods-1914.pdf, zuletzt aktualisiert am 20.03.2019, zuletzt geprüft am 20.03.2019.

BSI (2011): Leitfaden IT-Forensik. Hg. v. Bundesamt für Sicherheit in der Informationstechnik. 53133 Bonn. Online verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Leitfaden_IT-Forensik.pdf?__blob=publicationFile&v=2, zuletzt geprüft am 01.03.2019.

BSI (2019): Vorsorge für die IT-Forensik. DER.2.2. Hg. v. Bundesamt für Sicherheit in der Informationstechnik. Bonn. Online verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Download/FD_BS_Kompendium.pdf?__blob=publicationFile&v=6, zuletzt geprüft am 31.03.2019.

Card, S. K.; Mackinlay, J. (1997): The structure of the information visualization design space. In: Information Visualization, 1997 IEEE Symposium. VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium. Phoenix, AZ, USA, 20-21 Oct. 1997. Institute of Electrical and Electronics Engineers, Inc. Staff. Los Alamitos: IEEE Computer Society Press, S. 92–99.

Card, Stuart K.; Mackinlay, Jock D.; Shneiderman, Ben (Hg.) (1999): Readings in information visualization. Using vision to think. [Nachdr.]. San Francisco, Calif.: Morgan Kaufmann (The Morgan Kaufmann series in interactive technologies).

Carrier, Brian (2002): Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers. In: *International Journal of Digital Evidence* 1, S. 2003.

Carrier, Brian; Spafford, Eugene (2006): Categories of Digital Investigation Analysis Techniques Based On The Computer History Model. Hg. v. The Digital Forensic Research Conference. Lafayette, IN. Online verfügbar unter https://www.dfrws.org/sites/default/files/session-files/paper-

categories_of_digital_investigation_analysis_techniques_based_on_the_computer_history_model.pdf, zuletzt geprüft am 20.03.2018.

Casey, Eoghan (2011): Digital evidence and computer crime. Forensic science, computers and the Internet. 3rd ed. Amsterdam: Elsevier Acad. Press. Online verfügbar unter http://www.loc.gov/catdir/enhancements/fy1606/2010049562-d.html.

Chuvakin, Anton; Phillips, Chris; Schmidt, Kevin James; Moulder, Patricia (Hg.) (2013): Logging and log management. The authoritative guide to understanding the concepts surrounding logging and log management. Waltham, MA: Syngress. Online verfügbar unter http://proquest.tech.safaribooksonline.de/9781597496353.

CIRCL.lu (2019): Lookyloo. Hg. v. CIRCL. Computer Incident Response Center Luxembourg. Luxembourg. Online verfügbar unter https://github.com/CIRCL/lookyloo, zuletzt aktualisiert am 29.03.2019.

Cooper, Peter; Finley, Gail T.; Kaskenpalo, Petteri (2010): Towards standards in digital forensics education. In: Alison Clear (Hg.): Towards standards in digital forensics education. Proceedings of the 2010 ITiCSE working group reports. the 2010 ITiCSE working group reports. Ankara, Turkey. ACM Special Interest Group on Computer Science Education. New York, NY: ACM, S. 87.

Dewald, Andreas (2012): Formalisierung digitaler Spuren und ihre Einbettung in die Forensische Informatik. Online verfügbar unter https://pdfs.semanticscholar.org/5c5a/7cc3eada8f528606fc5a15d76063a3d3d530.pdf, zuletzt geprüft am 21.04.2019.

Dewald, Andreas; Freiling, Felix C.; Gruhn, Michael; Riess, Christian (Hg.) (2015): Forensische Informatik. 2. Auflage. Norderstedt: BoD - Books on Demand.

Digital Corpora (Hg.) (2018): Lone Wolf Scenario. Online verfügbar unter https://digitalcorpora.org/corpora/scenarios/2018-lone-wolf-scenario, zuletzt geprüft am 30.03.2019.

EDRM Duke Law (2019): Electronic Discovery Reference Model. Hg. v. Duke Law. Bolch Judicial Institute. Online verfügbar unter https://www.edrm.net/frameworks-and-standards/edrm-model/.

Elasticsearch (2019): Elasticsearch. The Elastic Stack. Online verfügbar unter https://www.elastic.co/de/products/elasticsearch, zuletzt geprüft am 25.03.2019.

Farmer, Dan (2000): What Are MACtimes? Online verfügbar unter http://www.drdobbs.com/what-are-mactimes/184404275, zuletzt geprüft am 12.04.2019.

Farmer, Dan; Venema, Wietse (2005): Forensic discovery. Boston: Addison-Wesley (Addison-Wesley professional computing series).

Fischer, Fabian (2016): Visual Analytics for Situational Awareness in Cyber Security. University Konstanz. Online verfügbar unter https://d-nb.info/1122561830/34, zuletzt geprüft am 12.03.2019.

Forensic Science Communications (Hg.) (2000): International Principles for Computer Evidence (Volume: 2. Issue: 2). Online verfügbar unter https://www.questia.com/library/journal/1G1-137863426/international-principles-for-computer-evidence, zuletzt geprüft am 02.04.2019.

Fowle, Ken; Schofield, Damian (2011): Visualising forensic data. Investigation to court.

Garfinkel, Simson L. (2010): Digital forensics research. The next 10 years. In: *Digital Investigation* 7, S64-S73. DOI: 10.1016/j.diin.2010.05.009.

Garfinkel, Simson L. (2013): Forensics Visualizations with Open Source Tools. Online verfügbar unter https://simson.net/ref/2013/2013-11-05_VizSec.pdf, zuletzt geprüft am 23.03.2019.

GeoTime (2019): GeoTime. Spot outliers. Hg. v. Uncharted. Online verfügbar unter https://geotime.com/wp-content/uploads/2016/01/slide-1-spotoutliers.png, zuletzt geprüft am 30.03.2019.

Gerson, Oliver Harry (2016): Das Recht auf Beschuldigung. Strafprozessuale Verfahrensbalance durch kommunikative Autonomie. Berlin, Boston: De Gruyter. Online verfügbar unter http://www.degruyter.com/doi/book/10.1515/9783110491517.

Geschonneck, Alexander; Kraus, Helmut; Schönfeldt, René (2014): Computer-Forensik. Computerstraftaten erkennen, ermitteln, aufklären. 6., aktualisierte und erweiterte

Auflage. Heidelberg, Germany: dpunkt.verlag (iX-Edition). Online verfügbar unter http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10852943.

Glatz, Eduard (2010): Visualizing host traffic through graphs. In: John Gerth (Hg.): Proceedings of the Seventh International Symposium on Visualization for Cyber Security. the Seventh International Symposium. Ottawa, Ontario, Canada. New York, NY: ACM, S. 58–63.

Gugelmann, David; Gasser, Fabian; Ager, Bernhard; Lenders, Vincent (2015): Hviz. HTTP(S) traffic aggregation and visualization for network forensics. In: *Digital Investigation* 12, S1-S11. DOI: 10.1016/j.diin.2015.01.005.

Hargreaves, Christopher James; Patterson, Jonathan (2012): An automated timeline reconstruction approach for digital forensic investigations. In: *Digital Investigation* 9, S69-S79. DOI: 10.1016/j.diin.2012.05.006.

Heer, Jeffrey; Card, Stuart K.; Landay, James A. (2005): Prefuse: A toolkit for interactive information visualization. In: Gerrit van der Veer (Hg.): Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. the SIGCHI conference. Portland, Oregon, USA, 2/4/2005 - 7/4/2005. ACM Special Interest Group on Computer-Human Interaction. New York, NY: ACM, S. 421.

Hendee, William R.; Wells, Peter N. T. (1997): The perception of visual information. 2. ed. New York, NY: Springer.

Hoelz, Bruno W. P.; Ralha, Célia G. (2013): A framework for semantic annotation of digital evidence. In: Sung Y. Shin und José Carlos Maldonado (Hg.): Applied computing 2013. The 28th Annual ACM Symposium on Applied Computing ; proceedings of the 2013 ACM Symposium on Applied Computing ; Coimbra, Portugal, March 18 - 22, 2013. the 28th Annual ACM Symposium. Coimbra, Portugal. Association for Computing Machinery; Annual ACM Symposium on Applied Computing; SAC. New York, NY: ACM.

Huang, Xiaodi; Eades, Peter; Lai, Wei (op. 2005): A framework of filtering, clustering and dynamic layout graphs for visualization. Sydney: Australian Computer Society

published in association with the ACM Digital Library (Australian computer science communications, v. 27 no. 1).

Hutchison, David; Fekete, Jean-Daniel; Kanade, Takeo; Kerren, Andreas; Kittler, Josef; Kleinberg, Jon M. et al. (Hg.) (2008): Information Visualization. Human-Centered Issues and Perspectives. Seminar on Information Visualization - Human-Centered Issues in Visual Representation, Interaction, and Evaluation. Berlin, Heidelberg: Springer (Lecture notes in computer science, 4950). Online verfügbar unter http://dx.doi.org/10.1007/978-3-540-70956-5.

Internet Engineering Task Force (2002): Guidelines for Evidence Collection and Archiving. RFC 3227. Online verfügbar unter https://www.ietf.org/rfc/rfc3227.txt, zuletzt geprüft am 20.04.2019.

Jankun-Kelly, T. J.; Stamps, Andrew S., Stamps; Wilson, David; Franck, Josh; Carver, Jeffery; Swan, J. Edward (2009): Visual Analysis for Textual Relationships in Digital Forensics Evidence. Online verfügbar unter http://web.cse.msstate.edu/~swan/publications/papers/2011_Jankun-Kelly-etal_Visual-Anal-Text_IV.pdf, zuletzt geprüft am 16.04.2019.

Katz, Evgeny; Halámek, Jan (Hg.) (2016): Forensic science. A multidisciplinary approach. Weinheim: Wiley-VCH Verlag. Online verfügbar unter http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1202056.

Kibana (2019): Kibana. The Elastic Stack. Online verfügbar unter https://www.elastic.co/de/products/kibana, zuletzt geprüft am 22.04.2019.

Krasser, S.; Conti, G.; Grizzard, J.; Gribschaw, J.; Owen, H. (2005): Real-time and forensic network data analysis using animated and coordinated visualization. In: Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, S. 42–49.

Labudde, Dirk; Spranger, Michael (2017): Forensik in der digitalen Welt. Moderne Methoden der forensischen Fallarbeit in der digitalen und digitalisierten realen Welt.

Berlin: Springer Spektrum. Online verfügbar unter http://dx.doi.org/10.1007/978-3-662-53801-2.

Lang Beebe, Nicole; Guynes Clark, Jan; Dietrich, Glenn B.; Ko, Myung S.; Ko, Daijin (2011): Post-retrieval search hit clustering to improve information retrieval effectiveness. Two digital forensics case studies. In: *Decision Support Systems* 51 (4), S. 732–744. DOI: 10.1016/j.dss.2011.01.009.

Locard, Edmond (1930): Die Kriminaluntersuchung und ihre wissenschaftlichen Methoden. 2. Aufl. Berlin: Kameradschaft Verlagsgesellschaft.

Logstash (2019): Logstash. The Elastic Stack. Online verfügbar unter https://www.elastic.co/de/products/logstash, zuletzt geprüft am 22.04.2018.

Lowman, Sarah G. (2010): Web History Visualisation for Forensic Investigations. Online verfügbar unter https://pdfs.semanticscholar.org/13af/b246855d527b633584b5aa59996d75ff456c.pdf?_ga=2.11652593.49619889.1561441576-762025258.1561441576, zuletzt geprüft am 04.05.2019.

Marty, Raffael (2009): Applied security visualization. Upper Saddle River, NJ: Addison-Wesley (Safari Tech Books Online). Online verfügbar unter http://proquest.tech.safaribooksonline.de/9780321585530.

McDermid, Val (2016): Anatomie des Verbrechens. Meilensteine der Forensik. Unter Mitarbeit von Doris Styron. München: Knaus A.

Meirelles, Isabel (2013): Design for information. An introduction to the histories, theories, and best practices behind effective information visualizations. s.l.: Rockport Publishers. Online verfügbar unter http://lib.myilibrary.com/detail.asp?id=524803.

Mihailowitsch, Fabian (2011): Don‘t scan, just ask. A new approach of identifying vulnerable web applications. Online verfügbar unter https://fahrplan.events.ccc.de/congress/2011/Fahrplan/attachments/2024_Dont_scan_just_ask_Fabian_Mihailowitsch.pdf, zuletzt geprüft am 05.04.2019.

Miller, David (2011): Security information and event management (SIEM) implementation. [enable real-time monitoring and analysis of security events ; respond quickly to attacks, log security data, and generate compliance reports ; get details on leading SIEM products - AlienVault OSSIM, Cisco-MARS, ArcSight ESM, and Q1 Labs QRadar]. New York, NY: McGraw-Hill (Network pro library).

Nordbø, André (2014): Data Visualization for Discovery of Digital Evidence in Email. Online verfügbar unter https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/198551, zuletzt geprüft am 25.03.2019.

Open Data Handbook (2019): http://opendatahandbook.org/guide/de/appendices/file-formats/. Online verfügbar unter http://opendatahandbook.org/guide/de/appendices/file-formats/, zuletzt geprüft am 21.04.2019.

Osborne, Grant; Thinyane, Hannah; Slay, Jill (2012): Visualizing Information in Digital Forensics, Bd. 383, S. 35–47.

Osborne, Grant; Turnbull, Benjamin (2009): Enhancing Computer Forensics Investigation through Visualisation and Data Exploitation. In: International Conference on Availability, Reliability and Security, 2009. ARES '09 ; 16 - 19 March 2009, Fukuoka, Japan ; [including workshop papers]. 2009 International Conference on Availability, Reliability and Security. Fukuoka, Japan. Fukuoka Kōgyō Daigaku; Secure Business Austria; International Conference on Availability, Reliability and Security; ARES - the International Dependability Conference. Piscataway, NJ: IEEE, S. 1012–1017.

Parasram, Shiva V. N. (2017): Digital forensics with Kali Linux. Perform data acquisition, digital investigation, and threat analysis using Kali Linux tools. Birmingham, UK: Packt Publishing. Online verfügbar unter http://proquest.tech.safaribooksonline.de/9781788625005.

Parker, Donn B. (1976): Crime by computer. New York, NY: Scribner.

Phan, Doantam (2007): Supporting the Visualization and Forensic Analysis of Network Events. Online verfügbar unter

http://graphics.stanford.edu/papers/dphan_thesis/doantam.phan.thesis.pdf, zuletzt geprüft am 20.03.2019.

Playfair, William (1786): The Commercial and Political Atlas. (edition 1801). Online verfügbar unter https://upload.wikimedia.org/wikipedia/commons/e/e0/1786_Playfair_-_Exports_and_Imports_of_Scotland_to_and_from_different_parts_for_one_Year_from_Christmas_1780_to_Christmas_1781.jpg, zuletzt geprüft am 01.05.2019.

Playfair, William; Wainer, Howard; Spence, Ian (2005): The commercial and political atlas and Statistical breviary. Cambridge: Cambridge University Press. Online verfügbar unter http://www.loc.gov/catdir/enhancements/fy0633/2005046971-d.html.

Pollitt, Mark (2010): A History of Digital Forensics. In: Kam-Pui Chow (Hg.): Advances in Digital Forensics VI. Sixth IFIP WG 11.9 International Conference on Digital Forensics, Hong Kong, China, January 4-6, 2010, Revised Selected Papers, Bd. 337. Berlin, Heidelberg: IFIP International Federation for Information Processing (IFIP Advances in Information and Communication Technology, 337), S. 3–15.

Sanders, Jan (2012): Introduction to Why-Because Analysis. Online verfügbar unter http://www.rvs.uni-bielefeld.de/research/WBA/WBA_Introduction.pdf, zuletzt geprüft am 07.04.2019.

Scheuermann, Gerik (2010): Informationsvisualisierung. Online verfügbar unter https://www.informatik.uni-leipzig.de/bsv/homepage/sites/default/files/Infovis_1-intro.pdf, zuletzt geprüft am 22.04.2019.

Schwartz, Moses; Liebrock, L. M. (2008): A Term Distribution Visualization Approach to Digital Forensic String Search. Online verfügbar unter https://vizsec.org/files/2008/Schwartz.pdf, zuletzt geprüft am 24.03.2019.

Shneiderman, Ben (1996): The eyes have it. A task by data type taxonomy for information visualizations. In: IN IEEE SYMPOSIUM ON VISUAL LANGUAGES, S. 336–343.

Staheli, Diane; Yu, Tamara; Crouser, R. Jordan; Damodaran, Suresh; Harrison, Lane; Nam, Kevin et al. (2014): Visualization Evaluation for Cyber Security: Trends and Future Directions. Online verfügbar unter

http://web.cs.wpi.edu/~ltharrison/docs/staheli2014visualization.pdf, zuletzt geprüft am 14.04.2019.

Stengel, Oliver; van Looy, Alexander; Wallaschkowski, Stephan (Hg.) (2017): Digitalzeitalter - Digitalgesellschaft. Das Ende des Industriezeitalters und der Beginn einer neuen Epoche. Springer-Verlag. Wiesbaden: Springer VS.

Stoll, Clifford (1989): The cuckoo's egg. Tracking a spy through the maze of computer espionage. 1. ed. New York: Doubleday.

Teelink, Sheldon; Erbacher, Robert F. (2006): Improving the computer forensic analysis process through visualization. In: *Commun. ACM* 49 (2), S. 71–75. DOI: 10.1145/1113034.1113073.

Tominski, Christian; Aigner, Wolfgang (2015): The TimeViz Browser. Online verfügbar unter https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html?, zuletzt geprüft am 20.03.2018.

Varga, Margaret; Winkelholz, Carsten; Träber-Burdin, Susan (2013): The Application of Visual Analytics to Cyber Security. Online verfügbar unter https://pdfs.semanticscholar.org/6068/847497e9dc2dd17fa655c0e32d955c3279b5.pdf, zuletzt geprüft am 13.03.2019.

Willer, Christoph (2012): PC-Forensik. [Daten suchen und wiederherstellen]. Dt. Orig.-Ausg., 1. Aufl. Böblingen: C & L Computer und Literaturverl. (Computer & Literatur).

Wundram, M.; Freiling, F. C.; Moch, C. (2013): Anti-forensics. The Next Step in Digital Forensics Tool Testing. In: 2013 Seventh International Conference on IT Security Incident Management and IT Forensics, S. 83–97.

Zimbardo, Philip G.; Gerrig, Richard J.; Graf, Ralf (2008): Psychologie. 18., aktual. Aufl. München: Pearson Studium (PS Psychologie). Online verfügbar unter http://www.myilibrary.com.

Zoss, Angela (2019): Data Visualization: Visualization Types. Online verfügbar unter https://guides.library.duke.edu/datavis/vis_types.

Thesen

- Nutzungsbarrieren, die aus einem hohen Einweisungs- und Ausbildungsaufwand resultieren, können durch den Einsatz von interaktiven Grafiken zur Datenanalyse reduziert werden.
- Es ist möglich, Ereignisse und Vorfälle zu klassifizieren, um deren Untersuchung zu vereinheitlichen.
- Methoden, Werkzeuge und Anwendungen, die während der forensischen Untersuchung eingesetzt werden, sollten frei verfügbar und erweiterbar sein.
- Anwendungen, die während der Datenanalyse eingesetzt werden, sollten es Analysten und Ermittlern ermöglichen, frei von systemspezifischen Eigenschaften der Quellsysteme zu agieren.
- Die Verwendung von interaktiven Grafiken während der Datenanalyse schließt eine abschließende Bewertung einzelner Nachweise nicht aus.
- Die Verwendung von interaktiven Grafiken zur Beschreibung und Dokumentation von Analyseergebnissen innerhalb der Abschlussdokumentation kann die Verständlichkeit erhöhen.
- Die Verwendungsmöglichkeiten von interaktiven Grafiken und Informationsgrafiken sind nicht auf die Datenanalyse beschränkt. Diese können vielmehr in allen Phasen der forensischen Untersuchung eingesetzt werden.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die hier vorliegende Arbeit selbstständig, ohne unerlaubte fremde Hilfe und nur unter Verwendung der in der Arbeit aufgeführten Hilfsmittel angefertigt habe.

Ort, Datum (Unterschrift)