MASTER KRIMINOLOGIE KRIMINALISTIK UND POLIZEIWISSENSCHAFT

JURISTISCHE FAKULTÄT • RUHR-UNIVERSITÄT BOCHUM

RUHR UNIVERSITÄT BOCHUM RUB

Big Data in Ermittlungsverfahren

– Grundlagen und Entwicklung eines Auswertungszyklus –

Masterarbeit

vorgelegt von

Kay Riedmüller

Abgabe: 05. Februar 2018

Erstgutachter: Prof. Dr. Jo Reichertz
Zweitgutachter: André Schulz

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

ANSI	American National Standards Institute
BAO	Besondere Aufbauorganisation
CRISP-DM	Cross-Industry Process for Data Mining
ETL	Extraktion, Transformation, Laden
GIS	Geografisches Informationssystem
GPS	Global Position System
HTML	HyperText Markup Language
IMEI	International Mobile Equipment Identification
IMSI	International Mobile Subscriber Identity
kBDA	kriminalistischer Big Data-Auswertungszyklus
KDD	Knowledge Discovery in Databases
NER	Named Entity Recognition
NSU	Nationalsozialistischer Untergrund
OLAP	OnLine Analytical Processing- Anwendungen
OODA	Observe, Orient, Decide, Act
RAF	Rote Armee Fraktion
SNA	Social Network Analysis
TKÜ	Telekommunikationsüberwachung
UTF	Unicode Transformation Format
WKWI	Wissenschaftliche Kommission Wirtschaftsinformatik und Gesellschaft für Informatik e.V.
WGS 84	World Geodic System 1984
XML	Extensible Mark-up Language

1 Einleitung

> „Informationserhebung und -verarbeitung, also das Sammeln, Sichten, Vergleichen und Analysieren von tat- und täterbezogenen Informationen machen das Wesen der Verbrechensbekämpfung aus. Sie kann und muß auf der Basis rechtsstaatlicher Gesetze von den modernsten Hilfsmitteln der Kriminaltechnik, der Nachrichtentechnik und der elektronischen Datenverarbeitung unterstützt werden.“[1] (Horst Herold, Präsident des Bundeskriminalamtes von 1971 bis 1981)

1.1 Ausgangslage und Themendarstellung

Die strukturierte Sammlung und Verarbeitung von Daten stellt schon immer eine wesentliche Aufgabe der Strafverfolgungsbehörden dar.[2] Relevante Daten werden zur Erfüllung ihrer gesetzlich bestimmten Aufgabe durch verschiedene Eingriffsmaßnahmen erhoben, innerhalb und außerhalb der erhebenden Behörde ausgetauscht und in unterschiedlicher Weise für eine spätere Verwendung gespeichert.[3] Wurden einst Daten in Papierform gespeichert und ausgetauscht, so erfolgt zunehmend die Datenverarbeitung mittels Computersystemen.[4] Hierbei ist eine stetig wachsende Datenmenge zu bewältigen. Bereits 2014 nannte der damals amtierende Präsident des Bundeskriminalamtes, Jörg Ziercke, Big Data als eine der Herausforderungen, denen sich die Polizei bei der täglichen Sachbearbeitung stellen muss.[5] Dass es sich hierbei nicht nur um einen kurzfristigen Trend handelt, ist an der aktuellen strategischen Aufstellung der Polizei zum Thema Daten- und Informationsverarbeitung erkennbar.[6]

Die Daten- und Informationsverarbeitung kann jedoch nicht auf die Datenerhebung, -speicherung und den Datenaustausch beschränkt werden. Eine wesentliche Aufgabe der Datenverarbeitung der Strafverfolgungsbehörden ist die Datenauswertung, die neben einer intelligenten Datenselektion[7], auch die Generierung neuen Wissens für ein konkretes Ermittlungsverfahren ermöglichen soll. Dabei rückt auch die Auswertung von Big Data immer mehr in den Fokus der Strafverfolgungsbehörden. So stellten Jo Reichertz und Sylvia Marlene Wilz[8] in einer qualitativen Studie zur Veränderung der polizeilichen Ermittlungsarbeit fest, dass die gezielte Auswertung von Big Data – als objektive Spur – immer mehr an Bedeutung gewinnt.

---

1 HENTSCHEL/PÖTZL 1986, S. 58.
2 Vgl. WESTPHAL 2010, S. 75.
3 Vgl. BT-Drs. 18/8596.
4 Vgl. JAEGER 2005, S. 415.
5 Vgl. ZIERCKE 2014, S. 14 f.
6 Vgl. hierzu MÜNCH 2017.
7 Vgl. ZIERCKE 2014, S. 15.
8 Vgl. REICHERTZ/WILZ 2016, S. 36 f.

Abseits dieser Domäne beschäftigt sich bereits seit längerer Zeit die Wirtschaft mit dem Thema der Datenauswertung. Hierdurch konnte sich der Wissenschaftszweig Wirtschaftsinformatik entwickeln und etablieren. So beschäftigt sich die Literatur zur Wirtschaftsinformatik intensiv mit der Auswertung von Daten einschließlich Big Data. Auch im Bereich der polizeilichen Arbeit werden seit einiger Zeit verschiedene Forschungen in diesem Themenfeld betrieben. Inhalt dieser Forschung ist vor allem die systematische Datenauswertung im Rahmen des Predictive Policing[9], welche die Verhinderung von Straftaten zum Ziel hat. Im Gegensatz hierzu gibt es insbesondere im deutschsprachigen Raum kaum wissenschaftliche Literatur, die sich mit der wissenschaftlich fundierten Auswertung von Big Data im Rahmen von Ermittlungsverfahren beschäftigt. Das betrifft nicht gleichermaßen alle Teilbereiche des breiten Spektrums der Ermittlungstätigkeit im Zusammenhang mit der Informationstechnik. Während das Feld der forensischen Untersuchung von Datensystemen in verschiedenen Publikationen thematisiert wird, ist das Thema der Auswertung großer strukturierter Datenbestände[10] in einem Ermittlungsverfahren, die aus eigenen und fremden Datenbanken gewonnen werden, in der Literatur bisher kaum thematisiert, obwohl in der öffentlichen Diskussion gerade dieses Thema im Rahmen der Vorratsdatenspeicherung öffentlich teilweise sehr leidenschaftlich geführt wird. Eine wissenschaftliche Beschäftigung mit dem Thema Auswertung von Big Data in Ermittlungsverfahren ist unumgänglich, treffen doch bei diesem Thema die neuen Möglichkeiten der Technik, die ethischen Vorstellungen der Bevölkerung und die Regularien des Rechtsstaates unweigerlich aufeinander.[11]

Deshalb beschäftigt sich diese Arbeit mit der Frage, wie Big Data in Ermittlungsverfahren unter Berücksichtigung der rechtstaatlichen Rahmenbedingungen ausgewertet werden können. Dazu werden alle relevanten technischen, rechtlichen und kriminalistischen Einflüsse, die sich aus der gesonderten Stellung der Big Data-Auswertung im Kontext eines Ermittlungsverfahrens ergeben, erarbeitet. Die Erkenntnisse hieraus fließen schließlich in einen systematisierten Prozess, der all diese Facetten beachtet.

Aufgrund des begrenzten Umfangs der Arbeit ist eine Eingrenzung der Betrachtung jedoch unumgänglich. Deshalb wird der Schwerpunkt auf der Ausgestaltung des Auswertungsprozesses und die Integration des Prozesses in Ermittlungsverfahren gelegt. Eine detaillierte Darstellung aller Handlungsschritte innerhalb der einzelnen Prozesspunkte kann aufgrund des begrenzten Umfangs nicht erfolgen.

---

[9] „Die Methoden, die unter dem Schlagwort ‚Predictive Policing' zusammengefasst werden, ermöglichen eine mathematisch-statistisch optimierte Nutzung von geocodierten und raumbezogenen Sozial- und Kriminaldaten, die dann die polizeiliche Präventions- und Ermittlungsarbeit substanziell unterstützen können." ROLFES 2017, S. 56.

[10] Zu den Begriffen mehr im Kapitel 2.

[11] Vgl. BITCOM 2014, S. 139 ff.

1.2 Methodik der Themenbearbeitung

Die Erarbeitung des Themas erfolgt mittels vorhandener wissenschaftlicher Literatur. Dazu dienen vor allem die Literatur der Wirtschaftsinformatik, die das Thema tangierenden Beiträge aus der Kriminalistik sowie die Literatur weiterer Bezugswissenschaften als Wissensquellen. Es werden entsprechende Fragen an verschiedene wissenschaftliche Literatur aus den Bereichen Informatik und Kriminalistik sowie anderen Bezugswissenschaften gestellt, um die verschiedenen Teilaspekte der Gesamtfragestellung zu beleuchten. Dabei stehen folgende Hypothesen im Vordergrund:

1. Der Begriff Big Data bezeichnet eine komplexe Ausgangssituation der Datenverarbeitung, die mehr ist als nur das Vorliegen von massenhaften Daten. Diese können auch im Ermittlungsverfahren vorliegen.

2. Big Data in Ermittlungsverfahren kann mit Hilfe von Big Data-Analysetechniken ausgewertet werden.

3. Zur Auswertung von Big Data im Ermittlungsverfahren ist ein eigener systematischer Prozess erforderlich, um die besonderen Anforderungen des Ermittlungsverfahrens abzubilden.

Durch die Prüfung der Hypothesen erfolgt ein Transfer des Wissens von einer überwiegend auf wirtschaftliche Ziele ausgerichteten Beschreibung der Thematik auf eine kriminalistische Zielrichtung. Hierdurch werden die bereits bestehenden Erkenntnisse nach neuen Gesichtspunkten strukturiert und somit neues kriminalistisches Wissen generiert. Es wird insbesondere auf spezifische Besonderheiten im kriminalistischen Kontext eingegangen, die für die Bezugswissenschaften, wie z. B. Informatik, keine Relevanz besitzen.[12]

1.3 Einordnung des Themas in die Kriminalistik

Das Thema Auswertung von *Big Data in Ermittlungsverfahren* ist grundsätzlich der Kriminalistik, „als [...] Lehre von der Erforschung des Sachverhalts im Strafrecht."[13] zuzuordnen. Kriminalistik wird in der Literatur in der Regel in die Teildisziplinen Kriminalstrategie, Kriminaltaktik und Kriminaltechnik[14] sowie in weitere Zweige wie z. B. Kriminaldienstkunde, Kriminalmedizin und Kriminalpsychologie untergliedert.[15] Nach der Begriffsbestimmung von Robert Weihmann und Hinrich de Vries[16] sowie von Horst Clages könnte die Befassung der Datenverarbeitung im Ermittlungsverfahren der Kriminaltechnik zugeordnet werden. Nach Clages wird

[12] Vgl. auch ACKERMANN/KORISTKA et al. 2000, S. 733.
[13] VRIES 2010, S. 28.
[14] Vgl. KUBE/SCHREIBER 1992, S. 2 f.
[15] Vgl. WEIHMANN/VRIES 2014, S. 53 f.
[16] Vgl. WEIHMANN/VRIES 2014, S. 139.

> „[. . .][u]nter Kriminaltechnik (auch naturwissenschaftlich-technische Kriminalistik) [. . .] die Anwendung naturwissenschaftlicher Methoden und Erkenntnisse sowie der Einsatz technischer Mittel bei der Suche, Sicherung, Untersuchung und Auswertung von materiellen Spuren im Rahmen der Verbrechensbekämpfung verstanden.“[17]

Die Kriminaltechnik umfasst mithin viele Fachgebiete wie z. B. Ballistik, Daktyloskopie, Akustik etc. Hier wird auch – entgegen der gegenständlichen Einschränkung von Clages auf materielle Spuren – die kriminalistische Untersuchung von Hard- und Software eingeordnet.[18] Da in dieser Arbeit jedoch nicht die kriminalistische Untersuchung *von* Hard- und Software, sondern die Nutzung dieser zu strafprozessualen Zwecken im Vordergrund steht, liegt die Frage nahe, ob die informationstechnischen Auswertungs- und Ermittlungsmethoden nicht anders eingeordnet werden müssen.

Hier bietet sich eine Anlehnung an die Einordnung der Wirtschaftsinformatik an. „Aufgabe der Wirtschaftsinformatik ist die Entwicklung und Anwendung von Theorien, Konzepten, Modellen, Methoden und Werkzeugen für die Analyse, Gestaltung und Nutzung von Informationssystemen.“[19] Innerhalb der Wirtschaftsinformatik wird versucht die Informatik – die sich selbst als Wissenschaft mit der maschinellen Datenverarbeitung befasst[20] – und die Betriebswirtschaftslehre sinnvoll zu integrieren.[21] Zudem wird in diesem Rahmen neben den Informationssystemen selbst auch das betriebliche Umfeld, Arbeitsabläufe und Personen betrachtet, die von diesem Prozessen direkt oder indirekt betroffen sind.[22]

Im Falle von Datenverarbeitung in Ermittlungsverfahren wird die Informatik mit Methoden der Kriminalistik, wie z. B. Beweisführung und Arbeitsorganisation, aber auch mit anderen Disziplinen, wie die Wahrnehmungspsychologie, Ethik, Soziologie und datenspezifischem Recht, in Verbindung gebracht. Damit geht die wissenschaftliche Befassung mit Datenverarbeitung in Ermittlungsverfahren über die Methodik zur Suche, Sicherung und Untersuchung von (informationstechnischen) Spuren hinaus. Eine Einordnung dieses Themas in den Bereich Kriminaltechnik ist daher nicht angezeigt. Aufgrund der aktuellen Entwicklung und der Komplexität dieses Themenbereiches ist die Schaffung einer neuen Disziplin in der Kriminalistik erforderlich, die eben diesen Betrachtungsschwerpunkt hat. In Anlehnung an den Begriff Wirtschaftsinformatik bietet sich die Bezeichnung Kriminalinformatik an.

Der Kriminalinformatik soll die wissenschaftliche Betrachtung der informationstechnischen Datenverarbeitung im Bereich der Strafverfolgung obliegen. Aufgabe

[17] CLAGES 2017b, S. 7.
[18] Vgl. ACKERMANN/CLAGES/ROLL 2011, S. 25.
[19] WKWI 2007, S. 319.
[20] Vgl. ABTS/MÜLDER 2017, S. 1.
[21] Vgl. ABTS/MÜLDER 2017, S. 3.
[22] Vgl. HANSEN/MENDLING/NEUMANN 2015, S. 11.

der Kriminalinformatik ist damit die Entwicklung und Anwendung von Theorien, Modellen und Methoden zur Gestaltung und Nutzung von Informationssystemen im Rahmen von Verbrechensbekämpfung.[23] Hierzu zählt auch die Auswertung elektronischer Spuren und die Auswirkungen dieser Methoden auf Organisation und Taktik bei den Strafverfolgungsbehörden. Ob es sich bei Kriminalinformatik um eine eigenständige Wissenschaft – wie es auch die Wirtschaftsinformatik ist – handelt, die sich aus dem Wissen der Bezugswissenschaften Kriminalistik, Informatik, Soziologie etc. bedient oder „nur“ um eine gleichberechtigte Teildisziplin der Kriminalistik, neben der Kriminaltechnik, kann an dieser Stelle dahingestellt sein.

Der bisherigen Betrachtung folgend bietet sich die Einordnung des Themas der vorliegenden Arbeit in die Kriminalinformatik als Teil der Kriminalistik an.

1.4 Aufbau der Arbeit

Nachfolgend wird die konkrete Herangehensweise zur Beantwortung der Forschungsfrage beschrieben. Die vorgelegte Arbeit ist neben diesem einleitenden Kapitel, das neben der Herleitung des Themas, die Darstellung der Forschungsfrage sowie die Einordnung des Themas im Definitionssystem der Kriminalistik zum Inhalt hat, in vier weitere Teile untergliedert. Eingeleitet wird die inhaltliche Aufarbeitung mit dem Kapitel 2 – Grundlagen zum Verständnis von Big Data. Hier werden die begrifflichen Grundlagen zum Verständnis des Phänomens Big Data geschaffen. Da die Verarbeitung von Daten und Informationen zu Generierung von Wissen Kern der Betrachtung dieser Arbeit ist, werden zunächst die Begriffe Daten, Informationen und Wissen definiert und gegeneinander abgegrenzt. Dies erfolgt aus einer informationstechnischen und einer informationswissenschaftlichen Sicht. Dabei werden bereits erste Bezüge zur Kriminalistik hergestellt. Da keine einheitliche Definition des Begriffs Big Data existiert,[24] wird anschließend eine Definition für die Arbeit diskutiert und festgelegt. Um das Grundverständnis für das technische Thema Big Data abzurunden, schließt sich an die Begriffsbestimmung eine Darstellung der grundlegenden Big Data-Auswertungsmethoden an. Auch hier wird ein Bezug zu Auswertungserfordernissen im Ermittlungsverfahren hergestellt. Zudem wird in diesem Abschnitt die Frage beantwortet, welchen Wert eine Big Data-Auswertung in der Wirtschaft und schließlich in Strafverfahren haben kann. Hierbei wird insbesondere die Frage beleuchtet, ob informationstechnische Daten Spuren oder gar Beweise in einem Strafverfahren sein können.

Nach der grundlegenden Einführung in das Thema wird im dritten Abschnitt unter der Überschrift Auswertungsverständnis in Ermittlungsverfahren der Prozess

[23] Bei der Entwicklung einer Definition für Kriminalinformatik wurde Anleihe an den Ausführungen der Wissenschaflichen Kommission Wirtschaftsinformatik und Gesellschaft für Informatik e. V. genommen; vgl. hierzu WKWI 2007, S. 319.

[24] Vgl. DORSCHEL 2015, S. 6.

der Auswertung betrachtet. Dazu muss zunächst der Begriff der Analyse, Analytics und Auswertung definiert und voneinander abgegrenzt werden, um einen korrekten Gebrauch dieser unterschiedlichen Begrifflichkeiten zu gewährleisten. Für das Verständnis des Auswertungsprozesses im Ermittlungserfahren ist an dieser Stelle erforderlich, kurz die möglichen Kategorien von Ermittlungsdaten vorzustellen, die sich unter dem dargestellten Begriff Big Data subsumieren lassen. Bevor in diesem Abschnitt abschließend wesentliche und bereits existierende Auswertungsprozesse der Kriminalistik und Wirtschaft gegenübergestellt werden, werden verschiedene Grenzen und Handlungsrahmen erörtert. Dazu werden sowohl psychologische wie auch rechtliche Aspekte diskutiert. In diesem Kontext wird auch auf den Gesichtspunkt der Datenqualität und ihrer Auswirkung auf die Wertschöpfung eingegangen.

Im vierten Abschnitt wird der neu entwickelte kriminalistische Big Data-Auswertungsprozess dargestellt. Die Gründe für die Notwendigkeit der Entwicklung eines eigenständigen Big Data-Auswertungsprozesses für Ermittlungsdaten wird hierzu noch im dritten Kapitel aufgezeigt. Im vierten Kapitel erfolgt nun die Erläuterung des gesamten kriminalistischen Big Data-Auswertungsprozesses. Dabei wird auf die Hintergründe und die grundsätzliche Ausgestaltung der einzelnen Schritte eingegangen.

In einem abschließenden Fazit werden die erarbeiteten Erkenntnisse zusammengefasst und in Bezug zueinander dargestellt. Hier wird resümiert, ob und wie kriminalistische Datenbestände, die Big Data zugeordnet werden können, im Ermittlungsverfahren bearbeitet werden können.

2 Grundlagen zum Verständnis von Big Data

In der vorliegenden Arbeit wird die Verarbeitung von Big Data in Ermittlungsverfahren besprochen und diskutiert. Dieses Thema wird – wie aus der Einleitung ersichtlich – mit Begriffen wie Daten, Informationen und Wissen in Verbindung gebracht oder gar gleichgesetzt.

In diesem Kapitel werden daher die Begriffe Daten, Informationen und Wissen definiert und voneinander abgegrenzt. Anschließend wird dargestellt, was unter Big Data zu verstehen ist. Eine solche grundlegende Betrachtung ist erforderlich, um anschließend in einem weiterführenden Kapitel prüfen zu können, wie Big Data im kriminalistischen Kontext ausgewertet werden kann.

2.1 Abgrenzung der Begriffe Daten, Information und Wissen

Eine in der Literatur weit verbreitete Sichtweise auf Daten, Informationen und Wissen und deren Zusammenhänge untereinander ist die Vorstellung, dass diese hierarchisch aufeinander aufgebaut sind.[25] Ausgangspunkt und unterste Ebene dieser Hierarchie bilden **Daten**. Sie bestehen aus Zeichen, die aufgrund von Strukturierungsregeln in einen Zusammenhang gebracht werden.[26] So stellt eine Zeichenkette (= Strukturierungsregel) mit mehreren Zahlen (bspw. ‚491711234567') zunächst einmal nur ein Datum dar. Dabei sind Daten nicht auf schriftliche Erscheinungsformen beschränkt. Ebenfalls können Daten eine akustische oder bildliche Erscheinungsform aufweisen.[27]

Wird einem Datum ein Kontext hinzugefügt, wird aus dem Datum eine **Information**.[28] Erhält z. B. die im obigen Beispiel genannte Zeichenkette ‚491711234567' den Kontext ‚Rufnummer eines Mobiltelefons', so entsteht aus dieser Kombination eine Information. Der Schritt der Kontextualisierung ist dabei maßgeblich für die Art der Information, die der Nutzer erhält. So kann ein geänderter Kontext in Verbindung mit demselben Datum die daraus entstehende Information vollkommen verändern. Erst der Kontext definiert eine Zeichenkette ‚221177' als Rufnummer, einen Code zur Entsperrung eines Mobiltelefons oder als Geburtsdatum ohne Trennpunkte. Beim Schritt des Setzens des Kontextes wird dem eigentlichen Datum ein weiteres Datum hinzugefügt, welches das erstere Datum beschreibt. Diese Daten über Daten werden Metadaten genannt.[29] Im öffentlichen Diskurs – insbesondere im Rahmen der Vorratsdatenspeicherung – wird überwiegend die Ansicht vertreten, dass die Zuschreibung, was Datum und was Metadatum ist, starr und unveränderlich ist. So werden bei Verbindungsdaten die Rufnummern der Gesprächsteilnehmer und der Zeitpunkt des Gesprächs generell als Metadaten bezeichnet.[30] Dieser Ansicht kann jedoch nicht gefolgt werden. Die Feststellung, ob ein Datum ein Metadatum ist, ist von der konkreten Perspektive abhängig. So stellen z. B. die Gesprächsinhalte einer Telefonüberwachungsmaßnahme die (Inhalts-)Daten dar, da der Zweck der Maßnahme eben die Erlangung dieser Inhalte ist. Weitere Daten, die zusätzlich zum Gesprächsinhalt erhoben werden und damit über den eigentlichen Erhebungszweck hinausgehen, stellen Metadaten zum überwachten Telefongespräch dar, da sie die Inhaltsdaten zusätzlich beschreiben. Sollen jedoch die Verbindungsdaten an sich ausgewertet werden, so verändert sich die Perspektive auf diese Verbindungsdaten und aus den Metadaten werden die eigentlichen Daten, aus denen eine Information gewonnen werden soll. Dabei können sogar die

[25] Vgl. u. a. MAINZER 2014, S. 156.
[26] Vgl. KRCMAR 2015, S. 4.
[27] Vgl. MERTENS/BODENDORF et al. 2017, S. 36.
[28] Vgl. APEL/BEHME et al. 2015, S. 3.
[29] Vgl. BITCOM 2014, S. 97.
[30] Vgl. u. a. KREMPL 2017, S. 235.

Gesprächsinhalte als beschreibendes Datum (z. B. bestimmte Schlagworte) selbst wiederum zum Metadatum bei der Betrachtung der Verbindungsdaten werden.

Um im nächsten Schritt aus der Information **Wissen** zu generieren, ist eine Vernetzung von Informationen erforderlich.[31] Erst mit der Verknüpfung verschiedener Informationen, wie z. B. der Zeitpunkt der Tat und der Kenntnis, dass die im vorherigen Beispiel genannte Rufnummer von dem Opfer unmittelbar vor der Tat angerufen worden ist, kann aus diesem Informationsgeflecht Wissen hervorgehen. Dieses Wissen entsteht, wenn der Nutzer die Informationen in Kombination miteinander interpretiert. Dabei ist der Schritt von Information zum Wissen eine höchst individuelle Handlung, die vom jeweiligen Nutzer der Informationen abhängig ist, da sie in dessen Kopf stattfindet.[32] Dieses Wissen ist deshalb – im Gegensatz zu Daten und Informationen – immer an Personen gebunden, da es von ihnen konstruiert wird.[33] Aus diesem Grund kann Wissen nicht direkt in elektronische Datenverarbeitungssysteme gespeichert und übermittelt werden. Eine Vermittlung von Wissen ist nur über den indirekten Weg möglich. Dazu muss der Empfänger der Informationen nicht nur in die Lage versetzt werden, eine Information zur Kenntnis zu nehmen, sondern er muss auch in der Lage sein, diese in sein Wissensnetz einzubetten. Hat er hierfür nicht die Voraussetzung, müssen ihm weitere Informationen mitgeliefert werden, die ihm helfen, die neuen Informationen in sein bestehendes Wissen einzubetten. Dabei können diese einordnenden Informationen ebenfalls sehr komplex sein, so dass in besonders schwierigen Fällen der Empfänger erst Wissen erlangen muss (z. B. Wissen über forensische Auswertungsmethoden), um die eigentlichen Informationen zu fachlichem Wissen transformieren zu können.

Im optimalen Fall hat der Empfänger mit Hilfe der Informationen das gleiche Wissen und Verständnis erlangt wie der Sender der Information.

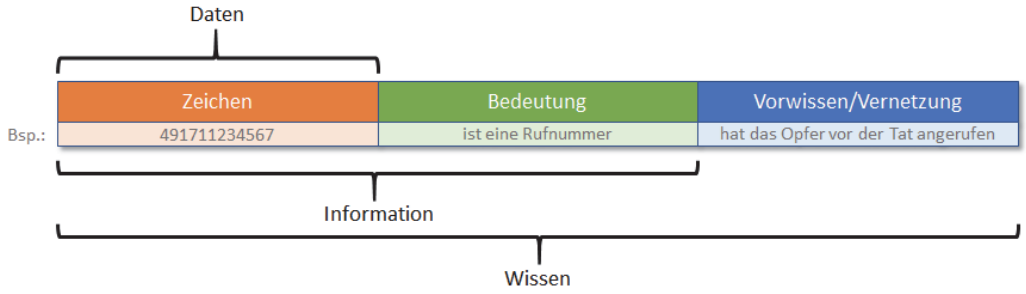

Abbildung 1: Zusammenhang zwischen Daten, Information und Wissen[34]

Nachdem nun die Begriffe Daten, Informationen und Wissen für diese Arbeit definiert wurden, wird im nächsten Abschnitt der Frage nachgegangen, wie sich

[31] Vgl. KRCMAR 2015, S. 4.
[32] Vgl. BURKHARD 2008, S. 304.
[33] Vgl. PROBST/RAUB/ROMHARDT 2012, S. 23.
[34] Abbildung 1: eigene Darstellung angelehnt an VÖLKER/SAUER/SIMON 2007, S. 60.

die eben diskutierten Abhängigkeiten zwischen Daten, Informationen und Wissen (siehe auch Abbildung 1) im Bereich von Big Data darstellen und welche Herausforderungen sich bei einer Auswertung von Big Data hierdurch ergeben.

2.2 Begriffsbestimmung Big Data

In der Bezeichnung Big Data ist der eben dargestellte Begriff des Datums enthalten. Wer daraus folgert, dass damit der Begriff als große Menge an Daten ausreichend definiert ist, verkennt, dass in diesem Kontext auch immer wieder von Informationen und Wissen gesprochen wird, welche aus Big Data gewonnen werden können. Um das Phänomen Big Data sachgerecht betrachten zu können, muss daher zunächst einmal beschrieben werden, was unter dem Begriff Big Data zu verstehen ist. Zudem soll an dieser Stelle geprüft werden, ob Big Data auch in Ermittlungsverfahren vorliegen kann, wie es z. B. vom ehemaligen BKA-Präsidenten Jörg Ziercke[35] dargestellt wurde. In der Literatur werden verschiedene Definitionen von Big Data angeführt.[36] Der Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (BITCOM)

> „[definiert] Big Data [...] [als] wirtschaftlich sinnvolle Gewinnung und Nutzung entscheidungsrelevanter Erkenntnisse aus qualitativ vielfältigen und unterschiedlich strukturierten Informationen, die einem schnellen Wandel unterliegen und in bisher ungekanntem Umfang anfallen.“[37]

Hier ist bereits erkennbar, dass an den Begriff Big Data umfangreiche Bedingungen geknüpft sind, die über eine Ansammlung von Daten hinausgehen. Die gängigen Definitionen von Big Data schreiben der Beschaffenheit sowie der Verarbeitung der Daten weitere wesentliche Charakteristika zu, die auch als die drei oder vier V's bezeichnet werden.[38] Dabei handelt es sich um die Merkmale Volume (Datenmenge), Velocity (Geschwindigkeit) und Variety (Vielfältigkeit). Als viertes Merkmal wird in diesem Zusammenhang zusätzlich Veracity (Glaubwürdigkeit) genannt.[39] Erhard Rahm et. al.[40] fügen diesen vier Merkmalen ein weiteres, fünftes Kennzeichen hinzu. Hierbei handelt es sich um das Kriterium Value (Werthaltigkeit). Andreas Gadatsch und Holm Landrock sehen neben den genannten Merkmalen auch „eine sehr hohe Anzahl an Nutzern von Berechnungsergebnissen“[41] als weiteres Kennzeichen für Big Data.

Aber sind diese Merkmale auch in den Datenbeständen der Kriminalpolizei zu finden? Um diese Frage beantworten zu können, werden nachfolgend die Big Data-

[35] Vgl. ZIERCKE 2014, S. 14.
[36] Vgl. DORSCHEL 2015, S. 6.
[37] BITCOM o. D.
[38] Vgl. DORSCHEL 2015, S. 6.
[39] Vgl. DORSCHEL 2015, S. 8.
[40] Vgl. RAHM/SAAKE/SATTLER 2015, S. 15.
[41] GADATSCH/LANDROCK 2017, S. 4.

Kriterien Volume, Velocity, Variety, Veracity und Value beschrieben und im kriminalistischen Kontext betrachtet.

2.2.1 Volume (Datenmenge)

Das bekannteste Kriterium für Big Data ist das Datenvolumen. Big Data ist demnach gekennzeichnet als das Vorhandensein gewaltiger Datenmengen und einer rasanten Zunahme dieser Datenmenge. Laut Aussage der Firma IBM wurden allein innerhalb der letzten zwei Jahren 90 % aller vorliegenden Daten generiert.[42] Dieses Wachstum steht dabei in einer direkten Korrelation zur steigenden Rechenleistung der informationstechnischen Datensysteme.[43]

Dabei nimmt nicht nur die absolute Anzahl der Datenquellen und der darin gespeicherten Datensätze zu (Datenbreite). Vor dem Hintergrund der ständig verbesserten Auswertungsmöglichkeiten steigt zudem innerhalb der Datenobjekte[44] auch die Anzahl der Attribute[45] (Datentiefe), wie z. B. Namen, Interessen, Aufenthaltsort, etc.[46] Große Datenmengen werden aber nicht nur durch die absolute Anzahl der Datensätze und der in diesen hinterlegten Attributen bestimmt. So kann bei hoher Spezialisierung des durch Daten abgebildeten Themas Big Data vorliegen, auch wenn das gemessene Datenvolumen selbst nicht groß ist.[47] In diesem Fall weicht die rein technische Betrachtung der Datengröße in Bytes einer fachlichen Betrachtung.

Das Aufkommen großer Datenmengen ist jedoch nicht nur auf die Wirtschaft begrenzt. Auch im privaten Bereich fallen umfangreiche Daten durch die Nutzung von privaten Computern, Internet und Telekommunikation an.[48] Die Datenmengen fließen schließlich bei gegebenem Anlass auch in die Datenverarbeitungsprozesse bei der kriminalpolizeilichen Sachbearbeitung ein.

Damit ist auch die Polizei bei ihrer Aufgabenerfüllung von großen Datenmengen betroffen. Dabei sind die Datenmengen vom jeweils zu ermittelnden Sachverhalt abhängig. Insbesondere in großen Ermittlungsverfahren, wie z. B. Ermittlungen zu den Hintergründen des Nationalsozialistischen Untergrundes (NSU), fielen große Datenmengen an. So wurden in diesem Verfahren für den Tatzeitraum Oktober bis November 2011 in Eisenach und Zwickau insgesamt mehr als vier Millionen Funkzellendaten erhoben.[49] Zudem wurden mehr als sieben Terabyte an Daten in

[42] Vgl. NEWMANN 2017.
[43] Vgl. DORSCHEL 2015, S. 7.
[44] „Objekte sind verbale, numerische oder visuelle Beschreibungen von Entitäten.“ HENNERMANN/WOLTERING 2014, S. 37. „Entitäten sind unterscheidbare Elemente der realen Welt oder der Vorstellungswelt.“ HENNERMANN/WOLTERING 2014, S. 36.
[45] „Attribute sind messbare oder benennbare Eigenschaften von Objekten.“ HENNERMANN/WOLTERING 2014, S. 38.
[46] Vgl. HIPP 2003, S. 1.
[47] Vgl. DORSCHEL 2015, S. 7.
[48] Vgl. SCHWARZ 2015, S. 315.
[49] Vgl. BT-Drs. 18/12950, S. 684.

unterschiedlichsten Datentöpfen, 7.000 Asservate[50], 1.500 Autokennzeichen und weitere umfangreiche Datensammlungen ausgewertet[51]. Diese nicht abschließende Aufzählung zeigt die enormen Datenmengen, die in einem einzelnen Ermittlungsverfahren anfallen können.

Solche Datenmengen sind nicht auf herausragende Ermittlungsverfahren beschränkt. Auch in regulären Ermittlungsverfahren, wie z. B. im Bereich der Wirtschaftskriminalität, können enorme Datenmengen anfallen. So wurden in einem Fall von Wirtschaftskriminalität, in dem wegen irregulärer Zahlungen von Provisionen gegen mehrere Personen ermittelt wurde, insgesamt 375 Asservate gesichert, die insgesamt 150 Millionen Dateien enthielten.[52] Generell sind die Verarbeitung von Datenvolumen im Bereich von Terabytes innerhalb eines Ermittlungsverfahrens üblich.[53]

Diese Datenmengen haben direkte Auswirkungen auf den Umgang mit ihnen. Während sich die Wirtschaft durch strategische Entscheidungen auf diese Datenmengen einstellen kann (z. B. durch gezielte Personalgewinnung und langfristige Umstellungen der Unternehmensprioritäten), besteht bei der kriminalistischen Datenerhebung die Herausforderung, dass die Datenmengen nicht zu jedem Zeitpunkt kontrolliert werden können. Vor allem zu Beginn von Ermittlungen ist oftmals nicht abschätzbar, welche Daten und Informationen für die Aufklärung des Falles relevant sind.[54] Das führt dazu, dass gerade zu Beginn der Ermittlungen eher mehr Daten – wie z. B. Funkzellendaten – erhoben werden, um diese vor Verlust durch Löschung zu schützen. Damit steht die Polizei mit gleichbleibender Personalstärke einer stark schwankenden Datenmenge gegenüber, die von Fall zu Fall unterschiedlich sein kann und somit in den Fällen von großen Datenmengen eine starke Belastung für das Personal bedeuten kann.

Die großen Datenmengen können sich dabei entweder auf die Menge unzähliger Personen, Sachen, etc. (Datenbreite) und/oder auf die Menge unzähliger Einzeldaten über eine Person, Sache, etc. (Datentiefe) beziehen. Damit stellt sich das Problem großer Datenmengen nicht nur der Wirtschaft, sondern auch der Polizei.

2.2.2 Velocity (Geschwindigkeit)

Das Big Data-Merkmal Velocity wird in der Literatur nicht einheitlich verstanden.[55] Eine Interpretation betrifft die Verarbeitungsgeschwindigkeit. Hiernach besteht aufgrund äußerer Umstände, wie z. B. Handlungsdruck durch die Konkurrenz die Notwendigkeit, die anfallenden großen Datenmengen auch schneller zu verarbeiten,

[50] Vgl. BT-Drs. 18/12950, S. 639.
[51] Vgl. BT-Drs. 18/12950, S. 885 ff.
[52] Vgl. HUFNAGEL/KOLLMANN 2015, S. 141.
[53] Vgl. HUFNAGEL/KOLLMANN 2015, S. 141; vgl. SPRANGER/LABUDDE 2017, S. 169.
[54] Vgl. BURGHARD 1993, S. 13 f.
[55] Vgl. DORSCHEL 2015, S. 7.

als dies in der Vergangenheit geschehen ist. Während früher Daten nach und nach angefallen sind und sukzessive ausgewertet werden konnten, ist heute aufgrund von Vernetzung und elektronischer Kommunikation sowie der damit einhergehenden ununterbrochenen Datenflut[56] ein anderer Umgang mit den Daten erforderlich, um im Wettbewerb bestehen zu können. Zudem wird aufgrund der ansteigenden Datenvolumina eine schnellere Datenverarbeitung erforderlich, damit Auswertungen den Entscheidungsträgern mindestens im gleichen Zeitraum wie zuvor vorliegen.[57]

In der Wirtschaft gilt es – um einen Wettbewerbsvorteil zu gewinnen – mehr und schneller Daten über Markt und Kunden zu sammeln und auszuwerten.[58] Dabei wird zunehmend die Verarbeitung von Daten in Echtzeit, also bereits unmittelbar nach Entstehung der Daten, erforderlich.[59] Die Auswertung von Datenströmen[60] wird hierbei immer bedeutender,[61] woraus wiederum neue technische Herausforderungen resultieren. So kann eine Auswertung der Informationen eines Datenstroms nicht in der Gesamtschau über alle Daten erfolgen, da es sich um dynamische, sich verändernde Daten handelt.

Vergleichbar zur Wirtschaft ist auch in Ermittlungsverfahren eine wichtige Anforderung eine hohe Verarbeitungsgeschwindigkeit der Daten. So kann eine durchschnittliche Auswertung im Bereich der Wirtschaftskriminalität leicht sechs Monate und mehr in Anspruch nehmen.[62] Aufgrund von Verjährungstatbeständen oder anderen Fristen (u. a. Haftprüfungstermine) müssen erhobene Daten, auch wenn sie in großen Mengen vorliegen, schnell ausgewertet werden.

Eine schnelle Datenverarbeitung ist vor allem dann erforderlich, wenn neben der Strafverfolgung auch gefahrenabwehrende Aspekte hinzukommen. Insbesondere im Bereich der organisierten Kriminalität und des Terrorismus müssen Daten schnell ausgewertet werden, um mögliche Gefahren für die Bevölkerung abzuwehren und das Strafverfahren zu sichern.[63] Auch bei Dauerstraftaten oder Tatserien, in denen Leib und Leben von Opfern anhaltend verletzt oder gefährdet werden, wie es z. B. im Bereich der Kinderpornografie gegeben ist, ist eine schnelle und sichere Datenauswertung unerlässlich.

---

[56] Vgl. HORVATH 2013, S. 1.
[57] Vgl. DORSCHEL 2015, S. 7.
[58] Vgl. BITCOM 2014, S. 13.
[59] Vgl. RAHM/SAAKE/SATTLER 2015, S. 14 f.
[60] „Ein Datenstrom (engl. Data Stream) ist eine geordnete, in der Länge meist unbeschränkte, Sequenz von x1, ... , xn Datenelementen, die in Echtzeit verarbeitet werden müssen, nicht wahlfrei zugreifbar sind und nur einmal oder in geringer Zahl erneut gelesen werden können.“ WEINER 2005, S. 3.
[61] Vgl. FASEL/MEIER 2016, S. 6.
[62] Vgl. SPRANGER/LABUDDE 2017, S. 169.
[63] Vgl. SPRANGER/LABUDDE 2017, S. 170.

2.2.3 Variety (Vielfältigkeit)

Ein weiteres Merkmal für Big Data stellt die Vielfältigkeit der Daten dar. Dieses Kriterium lässt sich in zwei Bereiche aufteilen. So unterscheiden sich die Daten erstens in der Art der Quellen (**Datenherkunft**), zweitens in der Form der Speicherung der Daten (**Datenabbildung**).

Getrieben vom Ziel, die individuellen Interessen des Kunden bei der Werbung und bei der Produktion einzubeziehen, ist die wirtschaftliche Auswertung einer Vielzahl von Informationen erforderlich, die weit über Selbstauskünfte des Kunden und einzelne Datenbankeinträge der Kundensysteme hinausgehen. Dieses wirkt sich nicht nur auf das Datenvolumen aus, sondern auch auf die Arten der Daten, die für eine solche Analyse herangezogen werden.[64] In die Analysen werden Daten und Datenquellen einbezogen, die bisher für wirtschaftliche Zwecke noch nicht kombiniert wurden.[65] Daten, die ursprünglich zu anderen Zwecken erhoben und gespeichert wurden, werden nun, im Rahmen einer Zweitverwertung[66], für die Auswertungen erschlossen.

Dieses Vorgehen wirkt sich direkt auf die Auswahl der Quellen, aber auch auf das Format der gemeinsam genutzten Daten aus. Nicht nur, dass die Quellen stark voneinander abweichen, es unterscheiden sich auch die Struktur der Daten teilweise erheblich.[67] Das liegt darin begründet, dass die Erhebung ursprünglich zu anderen Zwecken erfolgte und dabei keine vergleichbaren Erfassungserfordernisse vorlagen. Für wirtschaftliche Big Data-Analysen werden Daten unter anderem aus sozialen Netzwerken, Blogs, Tweets, E-Mails, Fotos und Videos, aber auch firmeninterne Daten, wie Sensordaten, die bei der Produktion anfallen, erhoben und genutzt.[68] Das führt dazu, dass gleichartige Daten in unterschiedlichen Formaten für die Auswertung vorliegen. So werden z. B. Kundendaten in entsprechenden Datenbanken sauber in den dafür vorgesehenen Datenfeldern für Name, Vorname, Straße und Hausnummer sowie Postleitzahl und Ort erfasst. Eine solche Erfassung ist jedoch bei verschiedenen Internetdiensten wie sozialen Netzwerken (z. B. Facebook) nicht zu erwarten, sodass – wenn überhaupt – nur mit Angaben zu Vor- und Nachnamen in einem Datenfeld sowie dem Wohnort in einem anderen Datenfeld gerechnet werden kann. Eine dezidierte und qualitätsgesicherte Erfassung wie sie für eine Kundendatenbank zur Erstellung von Rechnungen unabdingbar ist, wird für den Verwendungszweck eines sozialen Netzwerkes nicht erforderlich sein.

Die Herausforderungen ergeben sich aus der Notwendigkeit, diese heterogenen Daten für eine Auswertung zusammenzuführen, um sie anschließend in einem ge-

[64] Vgl. BITCOM 2014, S. 19.
[65] Vgl. HORVATH 2013, S. 2.
[66] Vgl. WEICHERT 2013, S. 133.
[67] Vgl. DORSCHEL 2015, S. 8.
[68] Vgl. HORVATH 2013, S. 1.

meinsamen Prozess auszuwerten.[69]

Strukturell lassen sich Daten dabei grob in die zwei Hauptgruppen „strukturierte Daten“ und „unstrukturierte Daten“ sowie in die Mischgruppe „semistrukturierte Daten“ einteilen. **Strukturiert** ist ein Datum, wenn nicht nur der Inhalt (z. B. ‚491711234567‘) sondern auch das strukturgebende Ordnungskriterium (z. B. ‚Telefonnummer‘) vorhanden ist.[70] An dieser Stelle kann das strukturierte Datum mit einer Information gleichgesetzt werden, da ein Kontext zu den Zeichen vorhanden ist. Diese Strukturinformationen werden überwiegend über eine Datenorganisation, wie z. B. in Form von Tabellen gewährleistet, in der die Tabellenüberschriften die Daten in der jeweiligen Spalte beschreiben. Es sind jedoch auch andere Formen der implizit strukturierten Daten denkbar, wie z. B. durch Tags[71] in XML[72]- oder HTML[73]-Dateien. Die Strukturinformation muss sich auf die fachlich kleinste Informationseinheit beziehen. Demnach wäre eine Adressliste strukturiert, wenn für die Einzelinformationen Land, Postleitzahl, Ort, Straße und Hausnummer jeweils eine eigene Struktureinheit besteht. Ein strukturiertes Datum würde nicht vorliegen, wenn die vollständig Adresse in einem einzigen Datenfeld oder einem Text und ohne eine weitere maschinell lesbare Systematik erfasst wäre. Können die Daten nicht direkt verarbeitet werden, weil sie keine Strukturelemente besitzen, werden diese Daten als **unstrukturierte** Daten bezeichnet.[74] Beispiel hierfür sind Fließtexte, Bilder und Videos. Für die Erschließung der Inhalte sind fortgeschrittene Techniken, wie z. B. Text Mining (siehe Kapitel 2.4.2), erforderlich. Enthält eine strukturierte Datei auch unstrukturierte Komponenten, die nicht durch standardisierte Instrumente verarbeitet werden können, so werden diese Daten als **semistrukturiert** bezeichnet.[75]

Relevant ist die Unterscheidung im Bereich Big Data deshalb, weil sich ca. 80 % bis 90 % aller relevanten Informationen eines Unternehmens in unstrukturierten Daten befindet.[76] Das bedeutet, dass diese Daten nicht direkt maschinell auswertbar sind. Um eine Auswertbarkeit aller Daten zu ermöglichen, müssen diese – wenn überhaupt möglich – in aufwendigen Prozessen erst strukturiert werden. Zum Verhältnis von strukturierten zu unstrukturierten Daten in polizeilichen Ermittlungsverfahren fehlen aktuelle Studien. Wie im Bereich der Wirtschaft werden auch

[69] Vgl. THEOBALD/FÖHL 2015, S. 118.

[70] Vgl. PIRO/GEBAUER 2015, S. 144.

[71] „Tags definieren [...] nicht nur die Darstellung des Inhalts, sondern können auch die Bedeutung des Inhalts erfassen (so könnte z.B. ein Tag <MAILADRESS> eine E-Mail-Adresse (E-Mail) einleiten“ LACKES/SIEPERMANN o. D.(c).

[72] „Abk. für Extensible Mark-up Language. XML ist ein [...][quasi] Standard zur Erstellung von strukturierten Dokumenten im World Wide Web oder in Intranets.“ LACKES/SIEPERMANN o. D.(c).

[73] „Abk. für HyperText Markup Language; [...] definierte Auszeichnungssprache, die die logischen Bestandteile eines Dokuments wie Überschriften und Aufzählungen beschreibt. HTML wird dazu benutzt, Dokumente für das World Wide Web zu erstellen, [...].“ LACKES/SIEPERMANN o. D.(a).

[74] Vgl. SEIDEL 2013, S. 4.

[75] Vgl. KING 2014, S. 35.

[76] Vgl. SEIDEL 2013, S. 4.

im privaten Umfeld unterschiedlich aufgebaute Daten an den unterschiedlichsten Orten erzeugt, wie z. B. PC, Handy, Internet-Router, Navigationssystem etc., und gespeichert.[77]

Im Rahmen von Sicherstellungen oder anderen Datenerhebungsmethoden greift die Polizei auf die eben beschriebenen Datenstrukturen zurück. Das trifft neben den strukturierten Daten auch auf unstrukturierte Daten zu. Zusätzlich können wichtige Informationen auch in nicht textbasierten Daten, wie z. B. in Bildern oder Videos im Bereich der Kinderpornografie, vorhanden sein.[78] Insgesamt sind Daten im Rahmen von Ermittlungsverfahren vielfältig. Abhängig vom jeweiligen Sachverhalt werden unterschiedlichste Datenbestände durch die verschiedenen Eingriffsmaßnahmen erhoben. So können ganze Datenbanken, private und geschäftliche Rechner, Datenbankauszüge von anderen Behörden und Betrieben, Videoaufnahmen, Verbindungsdaten von Mobilfunkbetreibern, etc. in den unterschiedlichsten Formaten zur kriminalpolizeilichen Bewertung und Verarbeitung vorliegen. Somit unterliegen die Daten, die im Rahmen polizeilicher Maßnahmen erhoben wurden, aufgrund ihrer Herkunft und Struktur ebenfalls der im Big Data typischen Vielfältigkeit.

2.2.4 Veracity (Glaubwürdigkeit)

Bei den eben diskutierten Merkmalen Volumen, Geschwindigkeit und Vielfältigkeit handelt es sich grundsätzlich um messbare technische Eigenschaften. Die Merkmale Glaubwürdigkeit und Werthaltigkeit stellen im Vergleich zu den erstgenannten aber zwei im Schwerpunkt fachlich-inhaltliche Eigenschaften der Daten dar.

Veracity beschreibt die Glaubwürdigkeit von Daten, die aus technischer Sicht durch die Datenqualität bestimmt wird. So haben Vollständigkeit und Fehlerfreiheit der Ausgangsdaten entscheidenden Einfluss auf die Glaubwürdigkeit der Auswertungsergebnisse.[79] Aus fachlich-inhaltlicher Sicht ergibt sich die Meinungspluralität der datenerzeugenden Menschen als weiterer wichtiger Faktor, der sich auf die Glaubwürdigkeit der Daten und Auswertungsergebnisse auswirkt. So haben z. B. Studien über die Nutzung des Web 2.0[80] gezeigt, dass nur bestimmte Gruppen aktiv Online-Beiträge zu bestimmten Themen verfassen.[81] Hierdurch kann eine Verzerrung des Meinungsbildes im Internet auftreten, die sich auf die Auswertungsergebnisse direkt auswirken kann.

---

[77] Vgl. SCHWARZ 2015, S. 304; vgl. HUFNAGEL/KOLLMANN 2015, S. 141.

[78] Vgl. SPRANGER/LABUDDE 2017, S. 172.

[79] Vgl. RAHM/SAAKE/SATTLER 2015, S. 15.

[80] Bei Web 2.0 handelt es sich um die Weiterentwicklung des Internets, „bei der nicht mehr die reine Verbreitung von Informationen bzw. der Produktverkauf durch Websitebetreiber, sondern die Beteiligung der Nutzer am Web und die Generierung weiteren Zusatznutzens im Vordergrund stehen." LACKES/SIEPERMANN o. D.(b).

[81] Vgl. THEOBALD/FÖHL 2015, S. 120.

Diesem Einfluss unterliegen auch die Daten, die im Rahmen von Ermittlungsverfahren erhoben werden. Auch diese können unterschiedliche Qualität aufweisen. Zudem besteht durch vorsätzliche Täuschung der Strafverfolgungsbehörden durch den Täter oder durch unbeabsichtigte Fehler der Datenzulieferer die Gefahr, dass sich technisch und fachlich fehlerhafte Daten auf das Auswertungsergebnis negativ auswirken. Aufgrund der komplexen Bearbeitungskette (siehe Kapitel 4) können sich Fehler über eine Art Stille-Post-Effekt weiter fortsetzen und verfestigen. Ein solcher Fehler ist im Ergebnis nicht leicht erkennbar. Daher ist im Kontext von Big Data Sorge zu tragen, dass eine Nachvollziehbarkeit und Begründbarkeit der Auswertungsergebnisse möglich ist.[82] Zudem ist stets im Vorfeld zu prüfen, ob die Daten aufgrund der Glaubwürdigkeit für die Nutzung in einem entsprechenden Anwendungsszenario geeignet sind.[83]

2.2.5 Value (Werthaltigkeit)

Das Merkmal Value bezieht sich auf den Umstand, dass die Daten auch für den jeweiligen Zweck nutzbringend sein müssen. Dieses Kriterium umfasst nicht nur die Anforderung, dass bei dieser Analyse neue Erkenntnisse für die Auftraggeber generiert werden,[84] sondern auch, dass die Ergebnisse für den Anwender verständlich sein müssen. Dazu sind geeignete aussagekräftige Analyseverfahren und angemessene visuelle Nutzerschnittstellen bereitzustellen.[85] Das Kriterium Werthaltigkeit wird auch gelegentlich unter dem Begriff **Smart Data** subsumiert, der eine sinnvolle Datenverwendung betont.[86] Eine solche Voraussetzung verlangt ein planerisches Vorgehen bereits vor der Datenerhebung und Auswertung. In der Wirtschaft könnte diese Planung z. B. durch Erkenntnisse der Sozialwissenschaften unterstützt werden.[87] Im Bereich der Kriminalistik ist hierfür eine strukturierte Ermittlungsplanung, einhergehend mit einer systematischen Hypothesenbildung, erforderlich (siehe Kapitel 4.1).

2.2.6 Zwischenfazit

Die Ausführungen haben gezeigt, dass Big Data nicht nur auf die reine Menge von Daten beschränkt werden kann. In der Literatur haben sich darüber hinaus weitere Kriterien etabliert, die Big Data definieren. Diese Kriterien können ebenfalls bei Datenbeständen im kriminalistischen Umfeld erfüllt sein, sodass in diesen Fällen von kriminalistischer Big Data gesprochen werden kann.

---

[82] Vgl. KING 2014, S. 35.
[83] Vgl. MERTENS/BODENDORF et al. 2017, S. 58.
[84] Vgl. GADATSCH/LANDROCK 2017, S. 3.
[85] Vgl. RAHM/SAAKE/SATTLER 2015, S. 15.
[86] Vgl. GADATSCH/LANDROCK 2017, S. 2.
[87] Vgl. FÖHL/THEOBALD 2015, S. 132.

Um sich schließlich der Auswertungsmöglichkeiten dieser kriminalistischen Big Data-Bestände zuwenden zu können, wird zunächst dargestellt, welche Informationen und Wissen durch die Big Data-Auswertung außerhalb der Strafverfolgung generiert werden können, um schließlich zu diskutieren, welchen Wert eine Big Data-Auswertung für das Ermittlungsverfahren haben kann. Dazu muss zunächst geprüft werden, ob elektronische Daten überhaupt als Spuren oder gar als Beweis in einem Strafverfahren in Betracht kommen können.

2.3 Wertschöpfung im Rahmen von Big Data

Der Kern von Big Data ist nicht das bloße Sammeln der Daten, sondern deren Auswertung. Dabei werden aus den gesammelten Rohdaten[88] entscheidungsrelevante Erkenntnisse extrahiert und aufbereitet, die für das eigene Geschäftsfeld von Interesse sind.[89]

2.3.1 Wertschöpfung für die Wirtschaft

Ergebnisse von Analysen haben Auswirkungen auf bestehende Geschäftsmodelle. Zusätzlich eröffnen sie aber auch neue Tätigkeitsfelder für bestehende oder neue Firmen.[90] Innerhalb bestehender Geschäftsgebiete ermöglichen die Erkenntnisse aus Big Data-Auswertungen eine Optimierung der Produktion, des Vertriebs und der Entscheidungsfähigkeit der Verantwortlichen dieser Betriebe. Durch Angebote, die sich an den Interessen der einzelnen Kunden ausrichten, werden Streuverluste im Bereich des Marketings reduziert.[91] In diesem Fall profitiert nicht nur das Unternehmen durch eine Kostenreduktion, sondern auch der Kunde aufgrund der auf ihn zugeschnittenen Produkte. Diese Individualisierung schlägt bis zur Produktion durch, indem Produkte nach den Bedürfnissen eines Kunden erstellt werden können. All diese Maßnahmen beziehen sich einerseits auf die Ausgestaltung des Produktes, welches durch den Betrieb gefertigt und dem Kunden angeboten wird. Andererseits ermöglicht die Big Data-Analyse neben diesem produktorientierten Ansatz unter Zugriff und der Zusammenführung der vielfältigsten Daten auch eine personenorientierte Auswertung. Diese wird z. B. für Risikoanalysen genutzt. Beispielhaft wäre hier die Berechnung der Kreditwürdigkeit eines Kunden oder das Schadenseintrittsrisiko für Versicherungsunternehmen zu benennen.[92] Hierbei werden neben den Kundendaten auch Sozialdaten in die Auswertung mit ein-

[88] Als Rohdaten werden in dieser Arbeit die Daten verstanden, die noch unverarbeitet in ihrer ursprünglichen Form vorliegen (technische Betrachtungsperspektive). Bei der Begriffsbestimmung von Rohdaten werden mögliche soziologische oder psychologische Selektionsmechanismen (vgl. hierzu GRUTZPALK 2013, S. 22 f.) vor der Datenerhebung nicht berücksichtigt.

[89] Vgl. BITCOM 2014, S. 73.

[90] Vgl. BITCOM 2014, S. 13.

[91] Vgl. OMRI 2015, S. 104 f.

[92] Vgl. BITCOM 2014, S. 19.

bezogen.[93] Eine weitere wesentliche Zielrichtung der Big Data-Analyse ist die Unterstützung bei Entscheidungsprozessen und Optimierungsstrategien. Durch die Auswertung großer Datenbestände werden bisher nicht erkannte Muster sichtbar, deren Kenntnis realistischere Planungen von Ressourcen sowie Schaffung effizienterer Abläufe ermöglicht.[94] Diese werden auch durch Vorhersagen von Trends, ein typisches Erkenntnisziel der Big Data-Auswertung in der Wirtschaft, unterstützt, indem Daten aus der Vergangenheit in die Zukunft extrapoliert werden.[95] Big Data-Analysen beschränken sich jedoch nicht nur auf die Optimierung bereits bestehender Geschäftsmodelle. Ein weiteres höchst erstrebenswertes Ziel ist die Erschließung neuer Geschäftsmodelle sowie die Schaffung neuer Produkte. Ähnlich wie bei der Risikoanalyse werden hierzu eine Vielzahl von Informationen zusammengeführt, um z. B. den Bedarf an neuen Serviceleistungen zu erkennen und Prognosen für die Einführung neuer Güter zu erstellen.[96]

2.3.2 Wertschöpfung für das Strafverfahren

In analoger Anwendung der eben dargestellten Auswertungsziele von Big Data in der Wirtschaft lassen sich auch für Strafverfahren verschiedene Zielrichtungen der polizeilichen Big Data-Auswertung identifizieren. So dienen die Ergebnisse von Big Data-Auswertungen ebenfalls der Planung polizeilicher Maßnahmen. Dabei können die Maßnahmen grundsätzlich in zwei Einsatzfelder eingeteilt werden. Zum einen können die gewonnenen Ergebnisse zur Vorhersage von Straftaten – dem sogenannten Predictive Policing – genutzt werden, was aber in dieser Arbeit aufgrund der Schwerpunktsetzung nicht weiter betrachtet wird. Zum anderen kann die Auswertung genutzt werden, um vergangene oder anhaltende Straftaten aufzuklären (siehe Kapitel 3.1). Bei der strafprozessualen Auswertung der Daten liegt ein personenbezogener Ansatz vor, bei dem es gilt, unbekannte Personen zu identifizieren sowie bekannte Beschuldigte zu überführen.

Die Aufklärung eines kriminalpolizeilich relevanten Sachverhaltes erfolgt dabei über die Abarbeitung von Spuren und endet schließlich im Beweisen der Täterschaft eines oder mehrerer Beschuldigter vor Gericht. Bei der Verwendung von Erkenntnissen aus Big Data-Auswertungen stellt sich nun die Frage, welche Rolle Daten in einem Strafverfahren haben können.

Daten als Spur

Zunächst stellt sich die Frage, ob Daten generell als Spuren angesehen werden können und wenn ja, welcher Spurenart sie zuzuordnen sind. Robert Weihmann

[93] Vgl. BITCOM 2014, S. 57.
[94] Vgl. WEICHERT 2013, S. 133.
[95] Vgl. SCHULMEYER/CHRISTIAN 2015, S. 324.
[96] Vgl. GADATSCH/LANDROCK 2017, S. 15.

und Hinrich de Vries führen zum Begriff Spur aus, dass dieser

> „[aus] dem Altgermanischen [stammt] und [. . .] ‚Tritt' oder ‚Fußabdruck' [bedeutet]. [. . .] [Er] war ursprünglich ein Begriff aus der **Jägersprache** [Hervorhebung im Original; Anm. des Autors] und bezeichnete das Ansetzen des Hundes auf die Fährte. In der übertragenen Bedeutung meint es das ‚hinterlassene Zeichen'."[97]

Folgt man dieser Definition, so stellt auch ein Datum und eine daraus gewonnene Information eine Spur dar, wenn diese zur Verfolgung einer Person im Sinne einer Fährte dient. Bei der Auswertung von Daten, die im Zusammenhang mit einer Straftat erzeugt und dann am Tatort – auch am Tatort im weiteren Sinne[98] – zurückgelassen wurden, soll ebenfalls die Fährte aufgenommen werden, die schließlich zum Täter führt. Dabei ist es jedoch unerheblich, ob eine elektronische Spur auf den Täter direkt verweist oder zu neuen Spuren führt, denen die Ermittlungsperson nachgehen kann. Diese Daten können vom Täter z. B. im Datenbestand eines Mobilfunkbetreibers erzeugt werden, wenn dieser am Tatort telefoniert. Bei der kriminalistischen Auswertung von Daten ist jedoch nicht immer ein Tatortbezug erforderlich. So können z. B. für die Rasterfahndung Daten herangezogen (siehe Kapitel 4.5.2), die keinen Tatortbezug aufweisen und dennoch durch Auswertung dieser Daten eine Spur generieren.

In der Gesamtschau der Argumentation kann festgestellt werden, dass Daten als Spuren in einem Ermittlungsverfahren betrachtet werden können. An dieser Stelle muss jedoch das Verständnis für kriminalistische Spuren erweitert werden. Eine materielle Veränderung der Umwelt[99] oder ein Tatortbezug sind in diesem Zusammenhang nicht immer von Relevanz, sondern nur die in den Daten innewohnenden Informationen und ggf. der Entstehungs- und Speicherort der Rohdaten.

Daten als Beweis

Nachdem festgestellt wurde, dass Daten als Spuren angesehen werden können, ist als nächstes die Frage zu klären, ob Daten auch als Beweismittel in einem Strafverfahren eingebracht werden können.

> „Als Beweis werden Tatsachen oder Erfahrungssätze bezeichnet, die in der Hauptverhandlung zu der Überzeugung des Gerichts führen müssen, dass eine Behauptung wahr oder unwahr ist, dass ein bestimmtes

[97] WEIHMANN/VRIES 2014, S. 148 f.

[98] Der Tatort im weiteren Sinne ist eine „Örtlichkeiten, die mit dem Ereignis im Zusammenhang stehen. Dazu können solche Orte gehören wie der Vorbereitungsort, der Annäherungsweg des Täters an das Tatobjekt, die nähere und die weitere Umgebung der Tatörtlichkeit, der eigentliche engere Tatort, der Fundort des Opfers, der Fluchtweg des Täters, Fluchtmittel, Fluchtfahrzeuge, der Verbringungs- und Verbergungsort der Beute oder Versteckorte von Tatwerkzeug." ROLL 2017, S. 83.

[99] Vgl. WEIHMANN/VRIES 2014, S. 149.

> Geschehen ohne vernünftige Zweifel sich so und nicht anders zugetragen hat."[100]

Dabei muss der Beweis mittels prozessual zulässigen, lückenlosen, nachvollziehbaren und logischen Argumenten erbracht werden.[101] Eine absolute oder mathematische Sicherheit wird jedoch nicht verlangt. Es genügt, dass der Richter[102] zu einer persönlichen Überzeugung gelangt.[103] Die Auswertung der Daten und somit die Daten selbst erlangen also nur dann einen Beweiswert, wenn die Ergebnisse und die daraus resultierenden Schlüsse für das Gericht nachvollziehbar und überzeugend sind. Daher ist für ein Gericht eine Sichtbarmachung der Daten und der daraus generierten Informationen und Wissen im Hinblick auf den zu ermittelnden Sachverhalt erforderlich. Diese Sichtbarmachung kann durch Berichte, Visualisierungen und auch durch Aussagen von Zeugen und Sachverständigen erfolgen. Demnach können über diesen Weg auch Daten in die Hauptverhandlung als mögliche Beweismittel eingebracht werden. Innerhalb der Beweislehre wird dabei zwischen direkten und indirekten Beweisen unterschieden. Ein direkter Beweis liegt dann vor, wenn dieser unmittelbar auf den nachzuweisenden Sachverhalt hinweist[104] und der Richter somit das Vorhandensein eines Tatbestandmerkmals ohne weitere Tatsachen oder Dritte erkennen kann.[105] Im Gegensatz dazu ist ein Beweis indirekt, wenn nur unter Anwendung von Denkgesetzen oder Erfahrungssätzen auf eine Tatsache geschlossen werden kann. Ein indirekter Beweis wird auch Indiz genannt.[106] Erheblich für ein Indiz ist nicht die Indiztatsache selbst, „[. . .] sondern der daran anknüpfende weitere Denkprozeß, kraft dessen auf das Gegebensein der rechtserheblichen weiteren Tatsache geschlossen wird."[107]

Demnach stellt des Aufzeichnung eines Telefonats, in dem z. B. eine Person eine andere zu einer Straftat aufgefordert hat, einen direkten Beweis für eine Anstiftung dar. Im Gegensatz dazu ist das aufgefundene Verbindungsdatum in einem Funkzellendatenbestand nur ein Indiz dafür, dass sich eine Person an einem bestimmten Ort befunden hat. Nach dem Erfahrungssatz, dass eine Person sein Mobiltelefon bei sich trägt, ist indirekt auf den Aufenthaltsort des Besitzers und damit auf eine mögliche Täterschaft zu schließen.

In der Regel genügt es nicht, dass nur eine beweiserhebliche Tatsache vorliegt, sondern es ist ein Zusammenspiel einer Mehrzahl von Beweismitteln erforderlich.[108] Dabei können sich mehrere Indizien – die jedes für sich unter Umständen

[100] ROLL 2013, S. 364.
[101] Vgl. CLAGES/ACKERMANN 2017, S. 47.
[102] Ausschließlich aus Gründen der Lesbarkeit wurde im Text die männliche Form gewählt, nichtsdestoweniger beziehen sich die Angaben auf Angehörige aller Geschlechter.
[103] Vgl. ROLL 2013, S. 364.
[104] Vgl. CLAGES/ACKERMANN 2017, S. 54.
[105] Vgl. NACK 1999, S. 35.
[106] Vgl. CLAGES/ACKERMANN 2017, S. 54.
[107] Vgl. BGH, Urteil vom 17.02.1970 – III ZR 139/67, BGHZ 53, 245-264.
[108] Vgl. CLAGES/ACKERMANN 2017, S. 47.

nicht ausreichen würden, um fehlerfrei ein Urteil bilden zu können – gegenseitig verstärken. Somit kann der Tatrichter in die Lage versetzt werden, die für ein Urteil erforderliche persönliche Überzeugung zu gewinnen.[109] Dabei muss das Auftreten der einzelnen Indizien in der vorliegenden Kombination rechnerisch so einzigartig sein, dass das Vorliegen dieser Indizienkombination bei zwei verschiedenen Personen nach den Regeln der Wahrscheinlichkeit nahezu ausgeschlossen ist.[110]

Zusammenfassend lässt sich feststellen, dass Daten in einem Strafverfahren die Rolle einer Spur sowie eines Beweises einnehmen können. Dabei ist jedoch zu beachten, dass das elektronische Datum als Sachbeweis immer noch dem Personalbeweis in der Hauptverhandlung untergeordnet ist,[111] auch wenn der Sachbeweis immer bedeutsamer wird.[112] Somit kann die Auswertung von Big Data in einem Ermittlungsverfahren auch einen Mehrwert für die Strafverfolgung haben. Die Wertschöpfung aus einer Big Data-Auswertung bleibt damit nicht auf die Wirtschaft beschränkt. Welche Techniken zu Big Data-Auswertung genutzt werden können, wird im nächsten Abschnitt beschrieben.

2.4 Techniken der Big Data-Auswertung

Um die Ziele der Big Data-Auswertung zu erreichen, bedienen sich die Wirtschaft und zunehmend auch die Strafverfolgungsbehörden verschiedener Techniken. Nachfolgend werden die Wichtigsten davon vorgestellt.

2.4.1 Data Mining

Eine bekannte Analysetechnik für Big Data ist das Data Mining. Unter Data Mining wird das automatisierte Aufspüren von bisher unbekannter Zusammenhängen, Mustern und Trends in den Datenbeständen einer oder mehreren Datenbanken verstanden.[113] Es handelt sich dabei nicht um eine einzelne Methode, sondern um einen Sammelbegriff mehrerer Datenanalysemethoden.[114] Data Mining steht nicht für sich allein, sondern ist der Kernprozess des Knowledge Discovery in Databases (KDD).[115] Der KDD ergänzt Data Mining um die Unterstützungsprozesse zur Datenvor- und -aufbereitung.[116]

Das Vorgehen beim Data Mining ist datengetrieben, das heißt, dass Data Mining kein hypothesenbasiertes Verfahren ist. Die Ergebnisse werden von den Daten ausgehend explorativ und automatisiert generiert, ohne das durch den Anwen-

[109] Vgl. NEUHAUS/ARTKÄMPER 2014, S. 45.
[110] Vgl. NEUHAUS/ARTKÄMPER 2014, S. 47 ff.
[111] Vgl. BGH, Beschluss vom 17.10.1983 – GSSt 1/83, BGHSt 32, 115-130.
[112] Vgl. NACK 1999, S. 33.
[113] Vgl. HANSEN/MENDLING/NEUMANN 2015, S. 291.
[114] Vgl. DÜRR 2004, S. 2.
[115] Vgl. FARKISCH 2011, S. 101.
[116] Vgl. KLIMETZEK 2013, S. 13.

der im Vorfeld Hypothesen erstellt werden, die er durch die Analyse der Daten zu überprüfen versucht.[117] Eine Erstellung von Untersuchungshypothesen im Vorfeld der Analyse ist demnach für Data Mining nicht charakteristisch. Andreas Reuß und Hans-Joachim Zwiesler[118] führen hierzu aus, dass eine völlige Unabhängigkeit von Hypothesen nicht möglich sei, da bereits durch die Eingrenzung des Untersuchungsfeldes gewisse Untersuchungserwartungen formuliert würden. Daher sprechen sie von einer tendenziell hypothesenfreien Analysemethode. Dieser Ansicht kann gefolgt werden, da das menschliche Handeln, also auch die Datenanalyse, durch individuelle Selektion und Deutungsmuster bestimmt ist.[119] Die wichtigsten Methoden des Data Mining sind die Klassifikation, das Clustering und die Assoziationsanalyse.[120]

Bei der **Klassifikationsanalyse** werden Datenobjekte automatisiert vorgegebenen Klassen zugeordnet. Das Ziel der Analyse besteht darin, die Zusammenhänge zu entdecken, die für die Zuordnung der Datenobjekte zu der jeweiligen Klasse verantwortlich sind. Hierfür müssen durch den Anwender Trainingsdaten an den Data Mining-Algorithmus übergeben werden.[121] Als Objekte werden Entitäten – beispielsweise Bankkunden – verstanden, zu denen weitere Informationen als Merkmale (Attribute), wie z. B. eine Örtlichkeit als Wohnanschrift, Beruf sowie der Familienstand erfasst sind.[122] Anhand dieser Merkmale werden die Objekte im Sinne eines zu analysierenden Problems – z. B. Prognose zum Kreditrückzahlungsverhalten – in verschiedenen Klassen zusammengefasst. Dazu werden dem Algorithmus typische Beispiele an Datenobjekten übergeben, die stellvertretend für die einzelnen Klassen sind. Anhand dieser generiert der Algorithmus autonom entsprechende Hypothesen, wodurch sich diese Klassen unterscheiden. Hierfür können Methoden der künstlichen Intelligenz eingesetzt werden.[123]

Die Methode des **Clustering** ist eng mit der Klassifizierung verbunden. Bei dieser Technik werden ebenfalls Objekte aufgrund ähnlicher Merkmale gruppiert. Im Unterschied zur Klassifizierung sind beim Clustering die Klassen nicht bekannt. Die Trennung der Klassen erfolgt aufgrund der gegebenen Unterschiede der Objekte innerhalb der einzelnen Cluster.[124] Die Aufgabe besteht darin, möglichst ähnliche Objekte in einem Cluster zusammenzufassen.[125] Für die polizeiliche Arbeit könnten durch das Clustering z. B. Personennamen oder Adressinformationen zusammengeführt werden, um so eventuelle Zusammenhänge erkennen zu können.

---

[117] Vgl. FARKISCH 2011, S. 103.
[118] Vgl. REUSS/ZWIESLER 2005, S. 6 f.
[119] Vgl. CIVELLI 2010a, S. 667 ff.
[120] Vgl. FARKISCH 2011, S. 103.
[121] Vgl. PIAZZA 2010, S. 42.
[122] Vgl. auch Fn 44.
[123] Vgl. DÜRR 2004, S. 3 f.
[124] Vgl. FARKISCH 2011, S. 106.
[125] Vgl. PIAZZA 2010, S. 45.

Bei der **Assoziationsanalyse** werden Daten nach interessanten Abhängigkeiten zwischen den einzelnen Untersuchungsobjekten überprüft. Das Ergebnis können sogenannte Wenn-Dann-Regeln sein,[126] wie z. B. wenn ein Kunde ein Smartphone kauft, dann kauft er auch innerhalb der nächsten Woche noch eine Schutzfolie für das Display. Solche Analysen können zu individuellen Angeboten in Form von Werbung führen, aber auch komplexe Analysen von Abhängigkeiten und Sequenzen für die Aufdeckung krimineller Handlungen – wie z. B. Kreditkartenbetrug – unterstützen.[127]

Eine weitere spezielle Form des Data Minings ist die des **Spatial Data Mining**. Bei dieser Methode werden räumliche Daten nach Mustern, die sich aufgrund der räumlichen Nähe ergeben, (tendenziell) hypothesenfrei ausgewertet.[128]

Das große Potenzial von Data Mining ist die automatisierte Entdeckung von Zusammenhängen und Mustern innerhalb fremder oder eigener Datenbestände. In Ermittlungsverfahren könnte auf diese Weise ein erster Überblick über sichergestellte Datenbestände gewonnen werden. Anhand von Trainingsdaten bekannter Täter oder relevanter Ereignisse (wie z. B. rechtswidrige Kontotransaktionen) könnten automatisiert Arbeitshypothesen zu der datentechnischen Repräsentation in den Datenbeständen erstellt werden. Auch wäre es möglich, z. B. bei Betrugssachverhalten eine Hypothese für typische rechtswidrige Buchungshandlungen zu generieren, die dann im gesamten Datenbestand abgeklärt werden können.

Wichtig in diesem Zusammenhang ist jedoch, dass die so generierten Muster und Korrelationen zunächst als zufällig betrachtet werden, solange man die zugrundliegenden Zusammenhänge nicht erkennt.[129] Daher ist es wichtig, die Ergebnisse zu hinterfragen, bevor auf diese Ermittlungshandlungen gestützt werden. Eine adäquate Dokumentation des gesamten Auswertungsprozesses ist daher gerade im Rahmen eines Ermittlungsverfahrens unabdingbar (siehe Kapitel 4.7).

2.4.2 Text Mining

Neben strukturierten Informationen liegen 80 bis 90 % der Informationen in Unternehmen in unstrukturierter Form vor.[130] Hierzu zählen Bilder, Videos oder Musik, aber auch Texte, wie z. B. Berichte oder E-Mails, die in wesentlichen Teilen auf natürlicher Sprache basieren. Zur Auswertung dieser Texte kann Text Mining genutzt werden. Text Mining dient der Extraktion von Informationen, die in den Texten enthalten und dem Anwender noch nicht bekannt sind.[131] Wie bei der Methode des Data Mining wird die Analyse tendenziell hypothesenfrei durchgeführt. Der Un-

[126] Vgl. DÜRR 2004, S. 3.
[127] Vgl. FARKISCH 2011, S. 100.
[128] Vgl. KLIMETZEK 2013, S. 36.
[129] Vgl. MAINZER 2014, S. 21.
[130] Vgl. SEIDEL 2013, S. 4.
[131] Vgl. SEIDEL 2013, S. 1 f.

terschied zwischen Data und Text Mining besteht darin, dass zumindest bei Text Mining der ursprüngliche Verfasser Kenntnis vom Inhalt hat.[132]

Die wichtigsten Methoden des Text Mining sind das Document Clustering, Document Classification und Information Extraction. Die Methoden des **Document Clustering** und **Document Classification** sind vergleichbar mit den bereits beschriebenen Methoden Klassifizierung und Clustering von Daten im Rahmen des Data Mining. Die Klassifizierung und das Clustering erfolgt nach dem fachlichen Inhalt der Dokumente, wie z. B. Rechnung oder E-Mail zu einem bestimmten Thema. Eine weitere Aufgabe des Text Mining ist die **Information Extraction**, welche darin besteht, aus unstrukturierten oder semi-strukturierten Daten strukturierte Daten zu gewinnen.[133] Innerhalb der Information Extraction ist die Named Entity Recognition (NER) die wichtigste Methode. „Eine Named Entity ist ein Wort oder eine Reihe von Wörtern, die einen Gegenstand der Realität benennen."[134] Das können z. B. Namen von Personen, Orte, Organisationen, Rufnummern, E-Mail-Adresse, Zeitangaben usw. sein. Bei der NER werden diese Namen im Text erkannt und anschließend den entsprechenden vordefinierten Typ zugeordnet. Dazu stehen verschiedene Techniken zur Verfügung, die regelbasiert alle Textbestandteile darauf prüfen, ob sie einem vordefinierten Typen entsprechen.[135]

Eine einfache Art der Datenextraktion ist die Nutzung Regulärer Ausdrücke[136]. So beschreibt z. B. das Suchmuster „$[0-3][012].[01][0-9].(19)|(20)[0-9][0-9]$" jedes Datum zwischen den 01.01.1900 bis 31.12.2099. Mit weiterführenden Regeln könnten aus einem so extrahierten Datumswert weitere Informationen wie z. B. Geburtsdaten von Personen gewonnen werden, die dann zur Anreicherung anderer Daten genutzt werden könnten. Wie eine Extraktion praktisch aussehen könnte, wird in der Anlage A am Beispiel eines funktionsfähigen Quellcodes zur Datenextraktion (ohne Vorverarbeitung) von E-Mailadressen dargestellt.

2.4.3 Visual Analytics

Die letzte grundsätzliche Methode, die im Rahmen dieser Arbeit besprochen wird, ist die der Visual Analytics. Bei Visual Analytics handelt es sich nicht um eine eigenständige Untersuchungsmethode für Big Data. Sie ergänzt und unterstützt die bereits besprochenen Analysemethoden wie z. B. Data Mining. Dabei werden die automatisierten Analysetechniken mit der (interaktiven) Darstellung der Ergebnisse kombiniert, um ein schnelles Verstehen und Schlussfolgern zu ermöglichen. Die Visualisierung verbindet somit die Maschine mit dem Menschen. Die Maschine hat

[132] Vgl. FELDEN 2006, S. 303.
[133] Vgl. SEIDEL 2013, S. 47.
[134] SEIDEL 2013, S. 49.
[135] Vgl. SEIDEL 2013, S. 47.
[136] „Reguläre Ausdrücke [...] sind eine generelle Notation zur Beschreibung von Textmustern [...]." FRIEDL 2008, S. 1.

in dieser Symbiose die Aufgabe, die Daten zum Teil automatisch oder autonom zu analysieren. Der Mensch steuert seine einzigartige Fähigkeit bei, mittels seines Wissens und seiner Erfahrung die durch die Maschine generierten Zwischenergebnisse zu interpretieren und zu neuem Wissen zu transferieren. Dadurch entsteht ein Prozess, in dem Mensch und Maschine arbeitsteilig agieren.[137]

Zumeist erlaubt erst die Visualisierung, die für die Fragestellung relevanten Schlüsse zu ziehen. In verschiedenen Szenarien ist es überhaupt erst durch eine abstrakte Visualisierung möglich, den Untersuchungsgegenstand nach den zu untersuchenden Kriterien zu betrachten.[138] Eine Visualisierung dient in diesen Fällen nicht der Präsentation von Informationen und Wissen, sondern ist ein essentieller Bestandteil der Analyse.

Der Prozess der Visual Analytics ist ein iterativer Prozess im Zusammenspiel von Mensch und der automatischen Datenanalyse. Aufgrund der automatischen Berechnungen werden Diagramme, Karten oder andere Darstellungsformen erzeugt, die der Mensch zur Modellbildung und -anpassung nutzt. Die neuen Erkenntnisse werden dann als Parameter den automatisierten Berechnungen übergeben, um dann wiederum neue Ergebnisse zu generieren, die schließlich wieder durch den Menschen bewertet werden.[139]

Auch in Ermittlungsverfahren wird die Methode der Visual Analytics genutzt, um große Datenmengen schnell und iterativ zu analysieren. Dafür kommen bei den Strafverfolgungsbehörden verschiedene Softwareprodukte, wie z. B. IBM Analyst's Notebook[140], HumanIT InfoZoom[141] oder ESRI ArcGIS[142] zum Einsatz, die alle eine Interaktion mit den Analysten bei der Datenanalyse zulassen und sogar explizit vorsehen. Schließlich gilt es, dass der Analyst bei der Auswertung von Big Data die Gesamtübersicht behält und auf diese Weise zielgerichtet weiterführende Auswertungen auf der Basis der polizeilichen Hypothesenbildung durchführen kann.

2.4.4 Zwischenfazit

Im vorliegenden Abschnitt wurden verschiedene Analysemethoden für Big Data dargestellt und überblicksartig auf Anwendbarkeit im kriminalistischen Kontext diskutiert. So ist festzustellen, dass die Big Data Analysemethoden der Wirtschaft auf kriminalistische Sachverhalte anwendbar sind.

Das nächste Kapitel beschäftigt sich mit der Durchführung einer solchen Big Data-Auswertung – mit dem Fokus auf der Auswertung in Strafverfahren.

[137] Vgl. KOHLHAMMER/FROFF/WIENER 2016, S. 317.
[138] Vgl. REITERER/JETTER 2013, S. 192.
[139] Vgl. KOHLHAMMER/FROFF/WIENER 2016, S. 318.
[140] Vgl. EDER 2005, S. 389.
[141] Vgl. HABERBERGER/TALARCZYK 2007, S. 368.
[142] Vgl. ESRI 2008, S. 2 f.

3 Auswertungsverständnis in Ermittlungsverfahren

Nachdem die Begriffe Daten, Informationen und Wissen für diese Arbeit definiert und die Wort- und Sinnbedeutung von Big Data beschrieben sowie die Auswertungsziele von Big Data dargestellt und im Hinblick auf eine kriminalistische Verwendung diskutiert wurden, wird nun besprochen, welche Auswertungsprozesse in der Kriminalistik und in der Wirtschaft bereits existieren. An dieser Stelle wird zudem geprüft, ob und wie die existierenden Prozesse auf die Auswertung von kriminalistischen Big Data-Beständen angewandt werden können.

Um sich dieser Frage zu nähern, ist es erforderlich, zunächst einmal festzustellen, was unter Auswertung und Analyse zu verstehen ist.

3.1 Abgrenzung der Begriffe Auswertung und Analyse

Im Zusammenhang mit Big Data in der Wirtschaft und bei der Polizei werden immer wieder die Begriffe (polizeiliche) Auswertung[143], Analyse und Analytik gebraucht und zum Teil unsauber verwendet.[144] Um das Phänomen Big Data im Bereich der polizeilichen Auswertung korrekt einordnen zu können, müssen daher zunächst diese Begriffe konkretisiert werden.

Unter **Analyse** verstehen Carsten Lanquillon und Hauke Mallow

> „[...] eine systematische Untersuchung einer Sache. Durch Untergliederung oder Zerlegung des Untersuchungsgegenstands – bei der Datenanalyse also der Daten – in seine Bestandteile sollen z. B. Strukturen, Auffälligkeiten, Regelmäßigkeiten oder Zusammenhänge aufgedeckt werden.“[145]

Diese Betrachtungsweise erfolgt aus einer wirtschaftlichen, in die Zukunft gerichteten Perspektive, in der aufgrund von Analyseergebnissen Entscheidungen getroffen und Maßnahmen für wirtschaftliche Unternehmen abgeleitet werden.[146]

Rolf Ackermann, Horst Clages und Holger Roll hingegen beschreiben Analyse im kriminalistischen Zusammenhang. Nach ihrer Ansicht soll eine Analyse durch „[...][d]ie isolierte Betrachtung der einzelnen Bestandteile eines ganzheitlichen Sachverhaltes [...] zu neuen Erkenntnissen für die Tataufklärung führen.“[147] Auch Waldemar Burghard beschreibt Analyse als Teil der kriminalistischen Handlungslehre, in der ein Sachverhaltes durch Aufgliederung in seine Teilaspekte – „auch [...][durch] Rückgang von der Wirkung (Ergebnis) auf die Ursache“ – untersucht wird.[148]

[143] Vgl. REEZ 2007, S. 18.
[144] Vgl. LANQUILLON/MALLOW 2015a, S. 55.
[145] LANQUILLON/MALLOW 2015a, S. 55.
[146] Vgl. LANQUILLON/MALLOW 2015a, S. 55 ff.
[147] ACKERMANN/CLAGES/ROLL 2011, S. 163.
[148] Vgl. BURGHARD 1993, S. 22.

Während Lanquillon und Mallow eine gegenwärtige Sichtweise einnehmen und für die Analyse eine in die Zukunft gerichtete Wirkung sehen, hat für Ackermann et al. sowie für Burghard die Analyse den Zweck, in der Gegenwart vorhandene Wirkungen (Spuren), die durch in der Vergangenheit liegende Ereignisse erzeugt wurden, zu erklären. Dieser Unterschied ist anhand der unterschiedlichen Ziele von Wirtschaft und Kriminalpolizei (im Bereich der Strafverfolgung) zu erklären.

Wirtschaftsunternehmen nutzen die Analyse dazu, um sich zukünftig im Wettbewerb besser zu positionieren. Die Kriminalpolizei hingegen versucht, vergangene Ereignisse aufzuarbeiten. Da in der vorgelegten Arbeit betrachtet wird, wie Big Data zur Aufklärung von Straftaten genutzt werden kann, wird nachfolgend Analyse als eine auf die Vergangenheit ausgerichtete Tätigkeit verstanden.

Von der Analyse ist begrifflich die **Analytik** zu unterscheiden. Der Begriff Analytik (englisch: analytics) bezeichnet „[...] die Lehre oder Kunst des Analysierens also der Durchführung von Datenanalysen. So wird der Begriff [...] [Analytik] auch unmittelbar für die Menge aller Analysemethoden verwendet."[149] Da der Begriff Analytik oft auch die die Analyse unterstützenden Technologien und Prozesse umfasst,[150] werden hierunter auch alle Maßnahmen der Analysevorbereitung und -nachbereitung subsumiert. Damit ist die Analyse ein Teilprozess der Analytik, was dazu führt, dass beide Begriffe nicht synonym genutzt werden können.

Innerhalb der Polizei wird zudem der Begriff der polizeilichen **Auswertung** genutzt. Dieser wird unterteilt in eine operative und eine strategische Auswertung.[151] Im Rahmen der operativen Auswertung werden konkrete Vorgangsdaten und polizeiliche Erkenntnisse analysiert und bewertet mit dem Ziel polizeiliche Ermittlungen zu initiieren, zu begleiten oder zu unterstützen.[152] Sie ist in der Regel auf die Betrachtung bestimmter Straftaten, Täter sowie Tätergruppierungen ausgerichtet.[153] Durch die operative Auswertung sollen bestehende Informationslücken aufgespürt und daraus Handlungen abgeleitet werden, wie diese Lücken geschlossen werden können.[154] Die strategische Auswertung hingegen hat das Ziel, mittels „Analyse und Bewertung von Vorgängen oder Vorgangskomplexen [...], ‚Polizeiliche Führungsinformationen' und andere Entscheidungsgrundlagen zu erarbeiten."[155] Die aus der strategischen Auswertung gewonnenen Erkenntnisse sollen die Polizei bei einer strukturierten und besser fokussierten Vorgehensweise ihrer operativen Arbeit unterstützen.[156] Grundsätzlich erfolgt die strategische Aus-

[149] LANQUILLON/MALLOW 2015a, S. 55.
[150] Vgl. LANQUILLON/MALLOW 2015a, S. 55.
[151] Vgl. KÖRFFER 2014, S. 147.
[152] Vgl. KOCH/SCHMIDT, *Einsatzlehre der Polizei*, Band 1, S. 65; zitiert nach LANZINGER/KELLNER 2002, S. 17.
[153] Vgl. KÖRFFER 2014, S. 147.
[154] Vgl. NONNINGER 2002, S. 6.
[155] PDV 388 (BGS) neu, zitiert nach LANZINGER/KELLNER 2002, S. 17.
[156] Vgl. NONNINGER 2002, S. 6.

wertung ohne konkreten Personenbezug,[157] da in diesem Zusammenhang z. B. abstrakte Lagebilder, Bekämpfungskonzepte oder Phänomenologie erstellt werden können.[158]

Damit umfasst der Begriff Auswertung sowohl eine strategische, in die Zukunft gerichtete Komponente (strategische Auswertung), wie sie auch im Bereich der Wirtschaft zur Anwendung kommt, als auch eine meist in die Vergangenheit gerichtete Komponente, die sich an Einzelsachverhalten orientiert (operative Auswertung). Beide Ausrichtungen bezeichnen jedoch – vergleichbar mit Analytik – den gesamten Prozess mit allen Unterstützungsprozessen.

Somit handelt es sich bei der kriminalistischen Auswertung von Big Data um einen Teilbereich der operativen Auswertung, in der die in der Gegenwart vorliegenden elektronischen Spuren mit dem Ziel beleuchtet werden, Informationen über die Ursache ihrer Entstehung zu gewinnen und damit offene Informationslücken bezüglich einzelner Aspekte einer strafbaren Handlung zu gewinnen.

3.2 Kategorisierung der Daten in Ermittlungsverfahren

Die Arbeit der Kriminalpolizei ist geprägt durch die Erhebung, Verarbeitung und Dokumentation von Daten und Informationen. Auch operative Eingriffsmaßnahmen, seien sie noch so handwerklich, werden im Anschluss an ihre Durchführung für die Ermittlungsakte dokumentiert.[159] Demnach liegen im Ermittlungsverfahren eine Vielzahl von Daten und Informationen vor, die sich aufgrund ihrer Quellen, ihrer Beschaffenheit und ihrem Zweck strukturieren lassen.

Im Mittelpunkt dieser Betrachtung stehen die **Verfahrensdaten**, die die inhaltliche Aufbereitung des zugrundeliegenden Sachverhaltes betreffen. Diese umfassen alle Informationen zu Personen, Sachen, Institutionen und Ereignissen, die das zu untersuchende Ereignis beschreiben. Diese Daten und Informationen werden entweder durch aktive Erhebung aus anderen Quellen eingeholt oder durch die Polizei selbst wahrgenommen und erzeugt. Sie lassen sich dabei – klassifiziert nach der Art der Entstehung – in polizeifremde und polizeieigene Verfahrensdaten unterscheiden. Zur Verdeutlichung sind in der Tabelle 1 einige nicht abschließende Beispiele aufgeführt.

Die Gruppen der polizeifremden und polizeieigenen Verfahrensdaten unterscheiden sich nicht nur durch unterschiedliche Quellen, sondern auch durch die Möglichkeit, auf die Art der Speicherung, Datenqualität und Umfang direkt Einfluss zu nehmen. Charakteristisch für polizeifremde Datenbestände ist, dass diese Daten außerhalb der Ermittlungsbehörde generiert wurden und in den Ermittlungsdaten-

[157] Vgl. KÖRFFER 2014, S. 147.
[158] Vgl. AHLF 2002, S. 3.
[159] Vgl. CLAGES 2017a, S. 67.

bestand überführt werden müssen. Diese Daten können z. B. bei Zeugen, anderen Behörden oder Firmen sowie beim Täter selbst entstanden sein. Eine Einflussnahme auf die Datenentstehung und die damit einhergehende Daten- und Informationsqualität besteht nicht. Im Gegensatz dazu bestimmen die Ermittlungsbehörden bei ihren eigenen Datenbeständen den Umfang und Qualität selbst.

Tabelle 1: Arten der Verfahrensdaten; geordnet nach ihrer Herkunft

Polizeifremde Verfahrensdaten	Polizeieigene Verfahrensdaten
• Sicherstellung/Beschlagnahme von Datenträgern und IT-Systemen • Rasterfahndung • Aufzeichnung von Telekommunikationsverkehr • Erhebung von Verbindungs- und Funkzellendaten • Internetdaten	• Ringalarmfahndungsdaten[160] • Schleppnetzfahndung • Schriftliche Vermerke und Berichte • Daten der Fall- und Vorgangsbearbeitungssystem

Eine weitere Daten- und Informationsebene betrifft die Koordination und Planung des Ermittlungsverfahrens an sich, welche bei großen, komplexen Sachverhalten sowie großen Ermittlungsgruppen ebenfalls sehr große Dimensionen annehmen kann. Diese werden im Bericht des Parlamentarischen Untersuchungsausschusses zum Nationalsozialistischen Untergrund[161] und den Schilderungen der Polizeiführer der BAO[162] TRIO (Ermittlungsverfahren zum NSU) Otmar Soukup und Wolfgang Barten[163] dargestellt. So wurden z. B. allein an den Tatorten in Eisenach und Zwickau über 6.800 Asservate sichergestellt, deren Bearbeitung innerhalb der BAO organisiert und nachgehalten werden musste. Zudem umfasste die Sachakte über 250.000 Seiten.[164] Ein solches Informationsaufkommen kann nur mit Hilfe von elektronischen Informationssammlungen koordiniert werden, die sich vorrangig auf die Arbeit der Polizei beziehen und damit nur indirekt auf den Sachverhalt. Diese Daten können als **Verfahrensadministrationsdaten** bezeichnet werden. Dieser Bereich der Informationsverarbeitung soll in dieser Arbeit nicht weiter betrachtet werden. Hier könnten weitere kriminalistische und polizeiwissenschaftliche Untersuchungen zur Wirksamkeit aus kriminaltaktischer Sicht bzw. die Etablierung dieser neuen technischen Erfordernisse im Polizeialltag erkenntnisreich sein.

[160] Bei der Ringalarmfahndung werden ringförmig um den Tatort Durchfahrtskontrollstellen eingerichtet, um die Kennzeichen der passierenden Fahrzeuge zu erheben; vgl. BURGHARD 1993, S. 69 f.
[161] Vgl. BT-Drs. 18/12950.
[162] Abkürzung für Besondere Aufbauorganisation; vgl. SOUKUP/BARTEN 2013, S. 22.
[163] Vgl. SOUKUP/BARTEN 2013, S. 22 ff.
[164] Vgl. SOUKUP/BARTEN 2013, S. 23.

3.3 Grenzen der Auswertung

Die Auswertung kann nicht losgelöst von dem Umfeld, in welchem sie durchgeführt wird, betrachtet werden. In diesem Kapitel werden drei wesentliche Einflussfaktoren, die sich auf die Auswertung im Allgemeinen und die Auswertung von Big Data im Besonderen auswirken, besprochen. Dabei handelt es sich um psychologische Aspekte, die einen Einfluss auf den Analysten haben, die Datenqualität, die insbesondere im Kontext von Big Data eine große Rolle spielt und die datenschutzrechtlichen Aspekte, die insbesondere für staatliche Big Data-Auswertung von großer Bedeutung sind. Auf weitere Aspekte wie z. B. ethische oder gesellschaftliche Gesichtspunkte wird an dieser Stelle aufgrund des begrenzten Umfangs der Arbeit nicht eingegangen.

3.3.1 Psychologische Aspekte

Der Auswertungsprozess ist – wie sich im nächsten Kapitel noch zeigen wird – kein vollständig automatisierter Prozess. Immer wieder greifen Menschen durch Entscheidungen und verschiedene Handlungen, wie z. B. Datenselektionen, in diesen ein. Ferner werden die Ergebnisse aus diesem Prozess von anderen Personen genutzt. Diese Personen, sei es ein Kriminalist, ein Staatsanwalt, ein Richter oder ein Verteidiger, unterliegen Wahrnehmungsverzerrungen, die sich auf die geforderte Ausgewogenheit und Unvoreingenommenheit auswirken können. Dabei haben Wahrnehmungsverzerrungen insbesondere bei Analysten große Auswirkungen, da auf ihren Ergebnissen weitere teilweise tiefgreifende Grundrechtseingriffe vorgenommen werden.[165] Ebenso können durch Verzerrung verursachte falsche Ergebnisse das Strafverfahren in eine Richtung lenken, die sich eventuell zu einem späteren Zeitpunkt nicht mehr korrigieren lassen, weil z. B. be- oder entlastende Beweismittel nicht mehr erhoben werden können. Das kann fatale Auswirkungen für den Beschuldigten, aber auch für die Opfer von Straftaten haben.

Siegfried Swan ordnet diese Verzerrungen unter den Begriff der (nachrichtendienstlichen) Auswertungpsychologie ein. Als Gegenstand dieser Disziplin sieht Swan

> „[. . .] [a]ll diejenigen psychischen Prozesse, die bei der Sammlung, Bewertung, Analyse, Zusammenfassung und Weitergabe von Informationen eine Rolle spielen, einschließlich solcher Umgebungsfaktoren, die auf die psychischen Prozesse Auswirkungen haben und somit das Ergebnis der Arbeit beeinflussen [. . .].“[166]

Nach seiner Auffassung wirken sich diese Prozesse auf verschiedene Arbeitsbe-

[165] Vgl. CIVELLI 2010a, S. 665.
[166] SWAN 2003, S. 10.

reiche der betroffenen Personen aus und haben direkten Einfluss auf den Auswertungsprozess.[167] Diese Aspekte sollen nun nachfolgend besprochen werden.

Ursächlich für Verzerrungen ist die Tatsache, dass ein Mensch nur einen kleinen Teil der Umweltreize aufnehmen kann. Dabei werden durch das Unterbewusstsein diejenigen Informationen herausgefiltert, die für die jeweilige Situation erforderlich sind. Dabei hat der Mensch selbst nur einen bedingten Einfluss auf die Wahrnehmungsauswahl. So wird die Wahrnehmung durch Aspekte wie Farben, Kontraste, Bewegungen aber auch durch tief verankerte Triebe gelenkt. Auch neuartige Signale ziehen die Aufmerksamkeit des Betrachters auf sich. Dabei tendiert dieser dazu, neue Informationen an bereits bekannte Informationen und Wissen anzuknüpfen und zur Verringerung der Komplexität diese in Strukturen und Gruppen einzuordnen und sie schließlich zu kategorisieren.[168] Verzerrungen lassen sich in drei Kategorien einteilen. So unterliegt der Mensch erstens solchen Verzerrungen, die durch das **soziale Umfeld** bedingt sind. Der von einer Gruppe ausgehende Druck kann nämlich dazu führen, dass ein Analyst mit einer noch ungefestigten Auffassung sein Ergebnis einer abweichenden Gruppenmeinung anpasst.[169]

Die zweite Kategorie umfasst die Verzerrungen, die aufgrund des **bestehenden oder noch nicht bestehenden Wissenstandes** hervorgerufen werden. Aufgrund des Ankereffekts, in dem eine bereits bestehende Information als Ausgangspunkt für alle weiteren Einschätzungs- und Entscheidungsprozesse dient, können einmal getroffene Aussagen den weiteren Entscheidungsprozess stark beeinflussen. Problematisch ist dieser Effekt vor allem deshalb, weil ein Anker völlig willkürlich gesetzt worden sein kann.[170] Somit hat nicht nur bereits vorhandenes Wissen Einfluss auf die weiteren Denkprozesse, sondern auch der Zeitpunkt und die Reihenfolge, in der neue Information erhoben oder präsentiert werden. Des Weiteren können sich Informationslücken nachteilig auf die Bewertung der gesamten Daten auswirken, da Individuen dazu neigen, fehlende Angaben abzuwerten.[171] Ein weiterer Verzerrungsaspekt liegt darin begründet, dass ein Analyst in seiner durch seine eigene Sozialisation erlangten Erfahrungen Informationen nach seinen eigenen individuellen Bewertungskriterien beurteilt und deren Glaubwürdigkeit und Plausibilität daran bemisst.[172] Aufgrund einer durch die Sozialisierung bedingten Erwartungshaltung wird ein Wahrnehmungsschema gebildet, „das vorgibt, auf was geachtet werden muss, was wichtig ist und wie die Interpretation der Wahrnehmung erfolgen soll.“[173] Einem solchen Wahrnehmungsschema unterliegt auch der kriminalistische Analyst, der seine Vorerfahrungen aus anderen Fällen und seiner

[167] Vgl. SWAN 2003, S. 11.
[168] Vgl. CIVELLI 2010a, S. 667.
[169] Vgl. CIVELLI 2010b, S. 719.
[170] Vgl. CIVELLI 2010b, S. 719.
[171] Vgl. WAGNER/MICHAELI 2009, S. 28.
[172] Vgl. CIVELLI 2010a, S. 668, 2010b, S. 721.
[173] WAGNER/MICHAELI 2009, S. 26.

persönlichen Herkunft auch auf neue Fälle anwendet, was letztendlich zu einer verzerrten Wahrnehmung und Bewertung der Analyseergebnisse führen kann. Zudem werden aufgrund der bereits ausgewerteten Daten mentale Modelle gebildet, die, obwohl sie schnell aufgebaut werden, sehr schwer zu ändern sind.[174]

Die dritte Kategorie der Verzerrungen betrifft die kognitiven Aspekte des Analysten. Am gravierendsten ist dabei das Phänomen der **kognitiven Dissonanz**. Neue Informationen, die dem bisherigen Sachstand und Überzeugen eines Analysten widersprechen, sorgen bei diesem für Unbehagen. Dieser innere Konflikt wird dann unbewusst ausgeglichen, indem die neuen Informationen abgelehnt oder zumindest abgeschwächt werden. Des Weiteren nimmt ein Analyst verstärkt die Informationen wahr, die der bisherigen Informationslage entsprechen.[175] Auch kann die Wahrnehmung und Interpretation der Informationen unbewusst so verändert werden, dass sie zu den bestehenden Überzeugungen passen.[176] Diese Anpassungen werden dann aufgegeben, wenn eine Bestätigung der ursprünglichen Einstellung nicht mehr möglich ist. Erst dann erfolgt eine Veränderung der eigenen Vorstellungen. Zusätzlich gibt es starke Voreinstellungen, bei denen ein solcher Umdenkprozess nicht stattfindet. Dies ist häufig bspw. bei Vorurteilen oder Stereotypen der Fall.[177]

Die Kenntnis über die psychologischen Verzerrungsfaktoren führt dazu, dass der Auswertungsprozess so konzipiert und durchgeführt werden muss, dass diese Faktoren einen möglichst geringen Einfluss auf die Ergebnisse und deren Verwertung haben.

3.3.2 Daten- und Informationsqualität

Neben personenabhängigen psychologischen Einflussfaktoren stellt die Datenqualität ein personenunabhängiges Kriterium dar, welches sich ebenfalls auf den Prozess auswirken kann. Dieser Einflussfaktor ist bedingt durch das Big Data-Merkmal Veracity (siehe Kapitel 2.2.4).

Die Qualität der Ausgangsdaten ist für die Auswertung im kriminalistischen Kontext – wie auch im wirtschaftlichen Umfeld – ein erfolgskritischer Faktor. Fehlerhafte Daten führen im Verlauf eines Datenflusses zu nicht absehbaren Folgefehlern, wenn diese nicht bereits am Anfang des Datenverarbeitungsprozesses erkannt und behandelt werden.[178] Für die Industrie ergeben sich bereits jetzt schon hohe wirtschaftliche Schäden, weil die Datenqualität unterschiedliche Güte aufweist.[179]

Abseits von monetären Gefahren können fehlerhafte Auswertungen in einem

[174] Vgl. WAGNER/MICHAELI 2009, S. 26.
[175] Vgl. CIVELLI 2010b, S. 723.
[176] Vgl. WAGNER/MICHAELI 2009, S. 26.
[177] Vgl. MANGOLD 2008, S. 260.
[178] Vgl. APEL/BEHME et al. 2015, S. vii.
[179] Vgl. ECKERSON 2002, S. 3.

Strafverfahren, die auf qualitativ schlechten Daten beruhen, schwerwiegende Folgen für betroffene Personen haben. Wird der Verdacht aufgrund einer mangelhaften Auswertung auf eine falsche Person gelenkt, kann dies weitreichende Folgen für die Grundrechte des Betroffenen haben. In Gemengelagen, in denen neben der Strafverfolgung auch gefahrenabwehrende Aspekte beachtet werden müssen (z. B. Entführungslagen oder Terrorismusverfahren), besteht zudem die Gefahr, dass diese Gefahrenlagen nicht mehr rechtzeitig abgewehrt werden können. Zudem können Verzögerungen und das Nichterkennen weiterer Handlungsoptionen dazu führen, dass Spuren unwiederbringlich verloren gehen.

Aber was wird unter Datenqualität verstanden? Qualität wird definiert als „[...] Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Tätigkeit, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen“.[180] Diese Definition bezieht sich nicht explizit auf Daten. Dennoch kann sie auf Daten – als Produkt oder Rohstoff – angewendet werden.[181] Guilherme Morbey hebt in seiner Definition zusätzlich den Bestimmungszweck heraus, indem er Datenqualität als „[...] Erfüllungsgrad der Gesamtheit der Anforderungen an die für einen bestimmten Zweck benötigten Daten [...]“[182] definiert. Auch für Helmut Krcmar[183] ist für die Daten- und Informationsqualität die Eignung für den jeweiligen Einsatzzweck ausschlaggebend. Daher ergeben sich zwei Dimensionen für die Bewertung der Datenqualität. Die erste Dimension bezieht sich auf die Eigenschaften und Merkmale der Daten selbst. Die zweite Dimension reflektiert, unter Berücksichtigung des Verwendungszweckes, auf die in den Daten innewohnende Information. Aus diesen und weiteren ähnlich lautenden Definitionen werden in der Literatur verschiedene Kriterien abgeleitet, die die Qualität von Daten bestimmen. Nachfolgend werden die wichtigsten Qualitätsmerkmale für Daten und Informationen aufgezeigt.

Die ersten Gruppe von Kriterien bezieht sich auf den Aufbau der einzelnen Daten sowie die Darstellung der Datensätze in Abhängigkeit zueinander.

Das erste hier zu nennende Kriterium ist die **Vollständigkeit** der Daten. Vollständigkeit bezeichnet die vollständige Befüllung einer durch Daten repräsentierten Entität.[184] So kann z. B. die Entität Person dann vollständig sein, wenn mindestens Angaben zum Vor- und Familienname, zum Geburtsdatum und Geburtsort sowie zum Geschlecht vorliegen. Welche Angaben letztlich vorhanden sein müssten, um eine Vollständigkeit zu bejahen, ist vom angestrebten Verwendungszweck abhängig. So kann das Geschlecht in verschiedenen Anwendungsszenarien irrelevant sein, wodurch ein Fehlen dieser Information die Datenqualität nicht negativ be-

[180] DIN 55350-11.
[181] Vgl. APEL/BEHME et al. 2015, S. 6.
[182] MORBEY 2011, S. 16.
[183] Vgl. KRCMAR 2015, S. 30.
[184] Vgl. APEL/BEHME et al. 2015, S. 27.

einflusst. Des Weiteren beschreibt dieses Kriterium auch den Datenbestand einer Datenbank an sich, wenn z. B. alle Datenobjekte im Zuge des Datenaufbereitungsprozesses (siehe Kapitel 4.4) in der Zieldatenbank gespeichert sind.[185] In diesem Fall wird die Datensatzanzahl vor dem Vorbereitungsprozess mit der Anzahl danach verglichen. Wurden Datensätze nicht gänzlich überführt, obwohl keine fachlichen Gründe, wie z. B. Plausibilitätsverletzungen, vorlagen, ist der Datenbestand nicht vollständig.

Das zweite Kriterium **Genauigkeit** beschreibt, ob die Daten den geforderten Detailierungsgrad erfüllen.[186] So erfüllt z. B. eine dezimale Geopositionsangabe im WGS 84-System[187] mit nur einer Nachkommastelle nicht das Kriterium Genauigkeit, wenn für eine genaue Visualisierung mit einer GIS[188]-Software mindestens sechs Nachkommastellen benötigt werden.

Ein für Big Data immer wiederkehrendes Datenqualitätsproblem ist die **Einheitlichkeit** der Daten. Einheitlich sind die Daten, wenn sie fortlaufend gleich in den Anwendungen abgebildet sind.[189] Das Problem ist in der Vielzahl der Datenquellen im Big Data-Kontext begründet und betrifft die Wirtschaft und das Ermittlungsverfahren gleichermaßen, da unterschiedlichste Datenquellen im Auswertungsprozess zusammengeführt und analysiert werden. Abhilfe wird durch eine Transformation der Daten im Datenvorbereitungsprozess geschaffen (siehe Kapitel 4.4.2).

Durch Nutzung verschiedener Datenquellen kann es neben unterschiedlichen Datenstrukturen auch zu ungewollten Datenduplikaten – sogenannten Redundanzen – kommen. Sind diese Redundanzen nicht gewollt, stellt die **Wiederholungsfreiheit** ein weiteres Qualitätskriterium dar.[190] Diesen ungewollten Wiederholungen kann durch Datenfusion[191] entgegengewirkt werden.

Die nächste Gruppe von Datenqualitätskriterien bezieht sich – wie durch Morbey und Krcmar zur Daten- und Informationsqualität angeführt – auf den angestrebten Verwendungszweck der Daten. Ein wesentliches Bewertungskriterium ist in diesem Zusammenhang die **Relevanz** der Daten und der daraus gewonnenen Informationen. Dabei ist festzustellen, inwieweit das Informationsangebot mit dem Informationsbedarf des Nutzers übereinstimmt, damit dieser seine Aufgabe vollständig erfüllen kann.[192] An dieser Stelle wird deutlich, dass Daten, die in ihren Quellsystemen von hoher Qualität waren, weil sie dem Anwender die erforderlichen Informationen zur Verfügung stellen konnten, im Rahmen einer Big Data Analyse von minderer

[185] Vgl. APEL/BEHME et al. 2015, S. 27.
[186] Vgl. FARKISCH 2011, S. 65.
[187] Die Abkürzung *WGS 84* steht für *World Geodic System 1984*, welches der Positionsbestimmung des Satelliten des Global Position System (GPS) zugrunde liegt; vgl. HAKE/GRÜNREICH/MENG 2002, S. 51.
[188] Abkürzung für *Geografisches Informationssystem*; vgl. ESRI 2008, S. 2.
[189] Vgl. APEL/BEHME et al. 2015, S. 9.
[190] Vgl. MERTENS/WIECZORREK 2000, S. 176.
[191] Unter Datenfusion wird das „[. . .] Zusammenführen und Vervollständigen lückenhafter Daten [. . .]“ verstanden. ZWIRNER 2015, S. 102.
[192] Vgl. APEL/BEHME et al. 2015, S. 27.

Qualität sein können, da durch die Zweckänderung eine ganz andere Bedarfslage beim neuen Nutzer entstanden sein kann.[193] Diese Qualitätsminderung kann ebenso innerhalb eines Ermittlungsverfahrens auftreten, selbst wenn die Datenerfassung und Auswertung durch die gleichen Bearbeiter erfolgt. So können Daten ursprünglich zu einem ganz anderen Zweck erfasst worden sein, als sie letztlich ausgewertet werden. Beispielsweise kann eine Liste von Kfz-Kennzeichen, die für einen Datenabgleich erstellt wurde, nicht nach Fahrzeugtyp und Wagenfarbe ausgewertet werden, da diese relevanten Informationen nicht in der Liste enthalten sind. An diesem Beispiel wird deutlich, dass Relevanz und Vollständigkeit einen starken Bezug zu einander aufweisen. Eine vorausschauende Planung der Datenverarbeitungsprozesse kann eine durch Zweckänderung verursachte Datenqualitätsminderung zumindest reduzieren.

Ein weiteres Kriterium, welches mit der Relevanz in enger Verbindung steht, ist die Zeitnähe der Daten. Diese betrifft zum einem die **Aktualität** der Daten und zum anderen die Bereitstellung der Daten für den Anwender.[194] So haben Daten eine schlechte Datenqualität, wenn sie bei der Anlieferung, der Verarbeitung oder bei der Ergebnisgenerierung nicht mehr die Wirklichkeit widerspiegeln. Diese Einschränkung wird durch die Schnelllebigkeit (Velocity) von Big Data verschärft. Hubert Österle und Boris Otto verdeutlichten die Veränderung der Datenqualität durch Zeitverlauf als sie sagten: „[Die] Datenqualität ändert sich über die Zeit, weil die Daten lediglich ein Abbild der Wirklichkeit darstellen, sich diese Wirklichkeit aber verändert.“[195] Dieser Aussage folgend verlieren jegliche Daten mit der Zeit an Qualität. Der andere Aspekt dieses Qualitätskriteriums ist der Zeitpunkt der Datenbereitstellung, worunter auch die Bereitstellung der Auswertungsergebnisse fällt. Die nicht zeitgerechte Bereitstellung von Auswertungsergebnissen – verzögert durch hohes Datenvolumen und Datenvielfalt mit der damit verbundenen Notwendigkeit langwieriger Verarbeitungsprozesse – ist vor allem bei anhaltenden Straftaten, Gefahrenlagen oder Haftsachen ein gravierendes Qualitätsproblem. Dieses Problem besteht unabhängig davon, ob die Daten den aktuellen Zustand der Welt widerspiegeln. So können Daten auch dann den beabsichtigten Verwendungszweck nicht erfüllen, wenn sie und die daraus gewonnenen Ergebnisse zu spät vorliegen.

Des Weiteren müssen die Attributwerte eines Datensatzes mit denen der modellierten Entität in der realen Welt übereinstimmen. Erst wenn dieses Kennzeichen gegeben ist, spricht man von **Korrektheit** oder **Fehlerfreiheit** der Daten.[196] Farkisch führt hierzu aus, dass auch die Metadaten, also die beschreibenden Daten der Daten, zum Datenqualitätsmerkmal Korrektheit zählen.[197] Insbesondere semi-

[193] Vgl. LANQUILLON/MALLOW 2015c, S. 262.
[194] Vgl. APEL/BEHME et al. 2015, S. 25.
[195] OTTO/ÖSTERLE 2016, S. 31.
[196] Vgl. APEL/BEHME et al. 2015, S. 8.
[197] Vgl. FARKISCH 2011, S. 65.

und unstrukturierte Daten weisen eine geringere Datenqualität auf, weil bei diesen definitionsgemäß keine oder zumindest unzureichende Metadaten zum entsprechenden Datum vorliegen. Diese Situation besteht dabei unabhängig davon, ob die Daten den modellierten Sachverhalt aus der realen Welt korrekt widerspiegeln, da korrekte Metadaten bei der Datenverarbeitung und später bei der Interpretation der Ergebnisse vonnöten sind.

Ein weiteres Datenqualitätskriterium stellt die **Verständlichkeit** der Daten dar. Verständlich sind Daten dann, wenn sie mit der Begrifflichkeit und der Struktur mit den Vorstellungen des Benutzer übereinstimmen.[198] Verständlichkeit ist damit kein Merkmal, welches sich ausschließlich auf den Inhalt und die Metadaten bezieht. Bei der Bewertung dieses Kriteriums muss – zum entsprechenden Zeitpunkt – der Nutzer, sein Wissen und seine Aufgabe mit betrachtet werden. So können z. B. Daten für einen Analysten verständlich sein, aber für einen kriminalpolizeilichen Sachbearbeiter wegen fehlender Kenntnisse über komplexe Analysemethoden unverständlich. Daher kann es erforderlich sein, dass zusätzlich zu den Ergebnisdaten weitere erläuternde Metadaten generiert werden.[199] Diese könnten z. B. in einer Dokumentation niedergelegt werden und erhöhen in diesem Fall die Datenqualität.

Die dritte Kriteriengruppe bezieht sich auf das Datenverarbeitungssystem, in welchem die Daten vorgehalten werden. Die wesentlichen systemseitigen Daten- bzw. Informationsqualitätskriterien sind Integrität, Zuverlässigkeit und Verfügbarkeit der Daten. Daniel Fasel und Andreas Meier verstehen unter **Datenintegrität** „[. . .] die fehlerfreie und korrekte Speicherung der Daten sowie ihren Schutz vor Zerstörung, vor Verlust, vor unbefugtem Zugriff und Missbrauch.“[200] Dabei ist vor allem der Schutz vor Zerstörung der Daten von besonderer Bedeutung. Der Schutz beginnt beim ersten Empfang der Daten bei einer Strafverfolgungsbehörde durch entsprechende sichere und nachvollziehbarer Archivierung. Berücksichtigt man, dass neben den sachverhaltsbezogenen Inhaltsdaten auch die bei der Auswertung und Analyse anfallenden Metadaten (z. B. Programmiercode, Begründung für das Vorgehen, Mappingtabellen[201] usw.) auch für das gesamte Strafverfahren schützenswert sind, so müssen auch diese vor Zerstörung bewahrt werden.

Das nächste Kriterium betrifft die Verknüpfung von Inhaltsdaten und Metadaten. **Zuverlässigkeit** beschreibt das Vorhandensein notwendiger Metadaten, um zu einem späteren Zeitpunkt nachvollziehen zu können, welche Herkunft eine Information hat und wie die hierfür genutzten Daten transformiert wurden. Zuverlässigkeit beginnt mit der Erfassung der Daten und dauert bis zur Erstellung der Analyseer-

[198] Vgl. APEL/BEHME et al. 2015, S. 9.
[199] Vgl. HEINRICH/RIEDL/STELZER 2014, S. 288.
[200] FASEL/MEIER 2016, S. 11.
[201] Unter Mapping wird die Zuordnung von Quelldatenfelder zu semantisch äquivalenten Zieldatenfeld verstanden. (vgl. HILDEBRAND et al. 2015, S. 124); Bsp: Familienname (Quelle) $\longmapsto$ Nachname (Ziel).

gebnisse an.[202]

Neben den bereits beschriebenen Aspekten spielt auch die **Verfügbarkeit** eine entscheidende Rolle für die Informations- und Datenqualität. Nur wenn sichergestellt ist, dass ein Nutzer auf die Daten zu jeder Zeit und mit relativ geringem Aufwand zugreifen kann, ist dieses Qualitätsmerkmal erfüllt.[203] Verfügbar bezieht sich nach dieser Auffassung auch auf Informationen aus einem Analyseergebnis, wenn diese in ausreichender Form schriftlich dokumentiert sind. Auch in diesem Fall ist sichergestellt, dass die Informationen dem Anwender mit geringem zeitlichen und faktischen Aufwand zur Verfügung stehen.

3.3.3 Regelungen des Datenschutzes

Neben dem faktischen Rahmen der Big Data-Auswertung, der durch Datenqualitätsmerkmale gesteckt wird sowie den psychologischen Einschränkungen, denen die am Auswertungsprozess beteiligten Personen unterlegen sein können, stellt Datenschutz weitere Grenzen dar. Die Regelungen des deutschen Datenschutzes gelten, wenn durch den Umgang mit personenbezogenen Daten eines Einzelnen dessen Persönlichkeitsrechte beeinträchtigt werden können.[204] Der Schutz erstreckt sich – unabhängig der Wertigkeit und Sensibilität der Daten[205] – auf die Erhebung, Verarbeitung und Nutzung von personenbezogenen Daten durch öffentliche Stellen von Bund und Ländern sowie durch nicht-öffentliche Stellen, soweit diese Datenverarbeitungsanlagen verwenden.[206]

§ 3 Abs. 1 BDSG definiert den Begriff personenbezogene Daten als „[. . .] Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).“ Der Rechteinhaber ist, wie aus der Legaldefinition ersichtlich, jede natürliche Person. Der Schutz nach dem BDSG endet mit dem Tod des Rechteinhabers. Danach sind seine Rechte weiterhin durch spezialgesetzliche Regelungen – wie z. B. durch § 203 Abs. 4 StGB[207] oder § 35 Abs. 5 SGB I[208] – geschützt.[209]

Unter Einzelangaben zu persönlichen Verhältnissen werden Daten gefasst, die Auskunft über die Person selbst geben. Hierunter fallen z. B. Name, Anschrift, Fa-

[202] Vgl. APEL/BEHME et al. 2015, S. 26.
[203] Vgl. BITCOM 2014, S. 141.
[204] Vgl. § 1 Abs. 1 BGSG; Bundesdatenschutzgesetz (BDSG) in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 10 Absatz 2 des Gesetzes vom 31. Oktober 2017 (BGBl. I S. 3618) geändert worden ist.
[205] Vgl. SCHAAR 2002, S. 43 ff.
[206] Vgl. § 1 Abs. 2 BDSG.
[207] Strafgesetzbuch in der Fassung der Bekanntmachung vom 13. November 1998 (BGBl. I S. 3322), das zuletzt durch Artikel 1 des Gesetzes vom 30. Oktober 2017 (BGBl. I S. 3618) geändert worden ist.
[208] Das Erste Buch Sozialgesetzbuch – Allgemeiner Teil – (Artikel I des Gesetzes vom 11. Dezember 1975, BGBl. I S. 3015), das zuletzt durch Artikel 5 des Gesetzes vom 14. August 2017 (BGBl. I S. 3214) geändert worden ist.
[209] Vgl. DÄUBLER et al. 2014, S. 125.

milienstand, Geburtsdatum, Beruf, Gesundheitszustand oder Überzeugungen.[210] Unter sachlichen Verhältnissen werden alle Sachdaten verstanden, die Auskunft über die Identität, die Merkmale oder das Verhalten einer natürlichen Person geben. Dabei muss ein Ergebniskontext, ein Zweckkontext oder ein Inhaltskontext zwischen dem Sachdatum und der Person bestehen.[211] Eine Sonderstellung nehmen Geodaten ein. Durch deren Auswertung können Angaben über den Aufenthalt, Nutzungsbeziehungen (z. B. Wohnung) oder Eigentum erlangt werden. Dabei unterscheidet die Literatur, ob es sich hierbei um Punktdaten oder Flächendaten handelt, denn je größer die Fläche ist, die durch die Geodaten beschrieben wird, desto geringer wird der Personenbezug.[212]

Maßgeblich ist, ob Daten einer bestimmten oder bestimmbaren Person zugeordnet werden können. Bestimmbar ist eine Person dann, wenn die datenbesitzende Stelle Kenntnisse, Mittel und Möglichkeiten hat, ohne unverhältnismäßigen Aufwand einen Personenbezug zwischen den Daten und einer natürlichen Person herzustellen.[213] Das bedeutet, dass Daten im Besitz eines Nutzers Personenbezug aufweisen können, aber für einen anderen Nutzer diesen nicht haben, wenn es nämlich diesem nicht möglich ist, direkt oder unter Verwendung von Zusatzwissen diese Daten einer natürlichen Person zuzuordnen. In diesem Fall spricht man von einem relativen Personenbezug. Ein solcher Personenbezug besteht z. B. für die Polizei bei einer Rufnummer oder einem Kfz-Kennzeichen, aber auch bei einer IP-Adresse, die mit Hilfe der Angabe zum Verbindungszeitpunkt durch einen Internetprovider einer bestimmten Person zugeordnet werden kann.

In der Literatur wird zusätzlich der objektive Personenbezug von verschiedenen Autoren beschrieben. Wolfgang Däubler et al.[214] vertreten die Auffassung, dass es nicht darauf ankommt, dass ein Datenbesitzer die Zusatzinformationen, die zur Herstellung der Kette zwischen Datum und natürliche Person erforderlich sind, besitzt oder leicht erwerben kann. Für sie ist es ausreichend, dass es nicht völlig unrealistisch ist, dass eine andere Stelle eben diese notwendigen Informationen zur Verfügung stellt. Eine Bestimmbarkeit ist nach ihrer Auffassung nur dann ausgeschlossen, „wenn das Zusatzwissen auf wenige begrenzt bleibt und dessen Preisgabe nicht nur rechtlich, sondern auch technisch-organisatorisch ausgeschlossen wird.“[215]

Diese unterschiedlichen Ansichten können für die Bewertung der Daten in einem Strafverfahren vernachlässigt werden, da es bei der Erhebung und Auswertung von Big Data gerade darauf ankommt, einen Personenbezug herzustellen. Daher wer-

[210] Vgl. GOLA et al. 2015, S. 85.
[211] Vgl. DÄUBLER et al. 2014, S. 129.
[212] Vgl. DÄUBLER et al. 2014, S. 130.
[213] Vgl. GOLA et al. 2015, S. 86.
[214] Vgl. DÄUBLER et al. 2014, S. 127.
[215] DÄUBLER et al. 2014, S. 127.

den auch nur solche Daten erhoben und in die Auswertung einbezogen, die für eine Identifizierung einer bestimmten Person dienlich sind. Aus diesem Grund sind die Regelungen der Datenschutzgesetze grundsätzlich bei Big Data-Auswertungen in Strafverfahren anzuwenden. Es ist auch unerheblich, ob die Daten von den Betroffenen freiwillig in einem anderen Zusammenhang wie z. B. durch eigene Veröffentlichung im Internet zur Verfügung gestellt worden sind. Es kommt lediglich darauf an, dass für den Betroffenen eine besondere Gefahrenlage entsteht, weil seine Daten gezielt zusammengetragen und unter Verwendung weiterer Daten ausgewertet werden.[216] Aufgrund dessen ergeben sich aus dem Datenschutzgesetz verschiedene Einschränkungen beim Umgang mit diesen Daten.

Dabei entsteht im Kontext von Big Data ein Spannungsfeld zwischen den Merkmalen Variety und Volume einerseits und den Datenschutzregelungen, vor allem zur Zweckbindung, zur Datensparsamkeit und zu Betroffenenrechten andererseits.[217] Grundsätzlich unterliegen alle Daten dem Grundsatz der Zweckbindung des § 14 Abs. 1 BDSG. Hiernach dürfen Daten nur für den Zweck verwendet werden, für den sie erhoben wurden. Werden Daten für ein konkretes Ermittlungsverfahren erhoben, so können sie auch uneingeschränkt genutzt – also auch ausgewertet – werden. Anders verhält es sich, wenn die primäre Datenerhebung zu einem anderen Zweck erfolgte. Ein solcher Fall liegt z. B. vor, wenn die Daten zur Gefahrenabwehr erhoben wurden und nun für die Strafverfolgung genutzt werden sollen. In dieser Konstellation muss geprüft werden, ob eine Zweckänderung der Daten gemäß § 14 Abs. 2 Nr. 1 BDSG möglich ist. Hierfür ist eine Rechtsvorschrift erforderlich, die eine Zweckänderung explizit zulässt.[218] So muss bei einer Verwendung von Daten zur Gefahrenabwehr auf Grundlage eines Polizeigesetzes dieses auch die Verwendung in einem Strafverfahren zulassen. In diesem Zusammenhang kommt spätestens seit dem Urteil des Bundesverfassungsgerichts vom 20.04.2016 zum Bundeskriminalamtgesetz (BKAG) zusätzlich das Konstrukt der hypothetischen Datenneuerhebung zum Tragen. Nach diesem Konstrukt ist die Nutzung von zu anderen Zwecken erhobenen Daten dann zulässig, wenn die konkreten Informationen aufgrund der Schwere der zur untersuchenden Straftat im Verfahren hätte selbst erhoben werden dürfen.[219] Mit der Neugestaltung des Bundeskriminalamtgesetztes wird in § 12 BKAG (gültig ab 25.05.2018) die hypothetische Datenneuerhebung bereits gesetzlich geregelt werden.[220] Eine entsprechende Norm ist in der Strafprozessordnung (StPO) nicht enthalten. Die Datenübernahme in das Strafverfahren erfolgt noch über die Generalklausel § 163

[216] Vgl. KÖRFFER 2014, S. 150.
[217] Vgl. KEUPER/SCHMIDT/SCHOMANN 2014, S. 25.
[218] Vgl. GOLA et al. 2015, S. 338.
[219] Vgl. BVerfG, Urteil vom 20.04.2016 – 1 BvR 966/09, BVerfGE 141, 220-378.
[220] Vgl. BT-Drs. 18/11163, S. 17.

Abs. 1 StPO.[221]

Ein weiteres Konfliktfeld ist Datensparsamkeit gemäß § 3a BDSG. Demnach ist

> „[. . .][d]ie Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen [. . .] an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen.“[222]

Diese Regelung steht dem Legalitätsprinzip gegenüber, wonach durch Polizei und Staatsanwaltschaft „alle keinen Aufschub gestattenden Anordnungen zu treffen [sind], um die Verdunkelung der Sache zu verhüten.“[223] Eine Einschätzung, welche Daten tatsächlich zur Zielerreichung erhoben werden müssen, kann in der polizeilichen Praxis weder zu Beginn und auch vielfach im späteren Verlauf nicht vorgenommen werden.[224] Daher wird dieses Konfliktfeld immer bestehen bleiben.

Des Weiteren ergeben sich aus den Datenschutzgesetzen eine Vielzahl von Betroffenenrechten. So muss der Betroffene die Information erhalten, welche Stelle seine Daten in welchem Ausmaß und zu welchem Zweck gespeichert hat. Diese Kenntnis ist erforderlich, um ein Recht auf Widerspruch, Berichtigung, Sperrung und Löschung wahrnehmen zu können.[225] Die Benachrichtigungspflichten in Strafverfahren regelt bei verdeckten Maßnahmen wie z. B. Raster- und Netzfahndung, IMSI-Catcher oder Erhebung von Verbindungsdaten § 101 StPO. Die Anzahl der zu benachrichtigenden Personen richtet sich nach der zugrundeliegenden Eingriffsmaßnahme und erfolgt erst, wenn der Untersuchungszweck nicht gefährdet wird. Nach dieser Regelung sind aber nicht alle Personen durch die Strafverfolgungsbehörden zu benachrichtigen, deren Daten bei einer Big Data-Auswertung einbezogen wurden. Dennoch müssen die Informationen für die zu benachrichtigenden Personen vorgehalten werden, um einen umfassenden Rechtsschutz der Betroffenen zu gewährleisten, wenn diese ihr Auskunftsrecht nach § 19 BDSG ausüben möchten.

3.3.4 Zwischenfazit

Diese Ausführungen haben gezeigt, dass es Einschränkungen gibt, denen Auswertungen von Big Data ausgesetzt sind. Diese ergeben sich aus der Person des Analysten und des Nutzers und der Qualität der Daten. Zudem bilden Datenschutzregelungen einen rechtlichen Rahmen, in denen sich die Datenauswertung insbe-

[221] Vgl. KAY 2013, S. 496.
[222] § 3a BDSG.
[223] § 163 Abs. 1 StPO; Strafprozessordnung (StPO) in der Fassung der Bekanntmachung vom 7. April 1987 (BGBl. I S. 1074, 1319), die zuletzt durch Artikel 2 des Gesetzes vom 30. Oktober 2017 (BGBl. I S. 3618) geändert worden ist.
[224] Vgl. BURGHARD 1993, S. 13 f.
[225] Vgl. GEISELBERGER/MOORSTEDT 2013, S. 141.

sondere im Ermittlungsverfahren bewegen muss. Hieraus ergeben sich verschiedene Zwänge, die in einem Auswertungsprozess beachtet werden müssen.

Welche Prozesse es zur Datenauswertung gibt und ob diese allen in diesem Abschnitt diskutierten Bedingungen gerecht werden, wird im nächsten Kapitel untersucht.

3.4 Darstellung von Auswertungs- und Analyseprozessen

In der informationstechnischen und kriminalistischen Literatur werden verschiedene Vorgehensweisen zur Datenauswertung und -analyse vorgestellt und besprochen. Diese sind zumeist als Zyklen beschrieben und haben die Daten- und Informationsauswertung zum Inhalt. Diese Zyklen weisen, wie nachfolgend dargestellt, verschiedene Parallelen, aber auch Unterschiede auf. Es wird nun geprüft, ob und welche dieser Vorgehensweisen für die Auswertung kriminalpolizeilicher Big Data-Bestände nutzbar sind, oder ob ein eigener Big Data-Auswertungszyklus erforderlich ist, um den Erfordernissen von Big Data zum einen und den Rahmenbedingungen eines Ermittlungsverfahrens zum anderen zu genügen.

3.4.1 Kriminalistische Auswertungszyklen

Im kriminalistischen Bereich beschreibt Hans Walter[226] den Auswertungsprozess aller kriminalistisch erhobenen Daten als Kreislauf. Nach seiner Auffassung beginnt dieser mit der Gewinnung eines Verdachts und erstreckt sich über die Datenanalyse, die Hypothesenbildung und die Programmbestimmung bis hin zur Beschaffung fehlender Daten, um schließlich wieder den Verdacht zu nähren, zu ändern oder zu verringern.

In den USA wurde ebenfalls ein Auswertungszirkel, der Intelligence Cycle, entwickelt, der auch aus fünf Phasen besteht. Dieser unterscheidet sich in seiner Abfolge von Walters kriminalistischem Zyklus. Startpunkt dieses Prozesses ist die *planing and direction*-Phase, in der bestimmt wird, welche Daten erhoben werden sollen. Danach erfolgt die *collection*-Phase, also die Sammlung der vorhandenen Daten. Im dritten Schritt *processing and exploitation* werden die Daten verarbeitet und verwertet, damit diese im vierten Schritt *analysis and production* analysiert werden können. In der letzten Phase *dissemination and integration* werden die Daten verteilt und so in die bereits vorhandenen Daten integriert, sodass die Planung neu ausgerichtet werden und der Prozess von vorne beginnen kann.[227]

Die Autoren Büchler, Meywirth, Kalseher und Vogt[228] trugen in ihrem Aufsatz

[226] Vgl. WALDER/HANSJAKOB 2016, S. 93.
[227] Vgl. WALDER/HANSJAKOB 2016, S. 92 f.
[228] Vgl. BÜCHLER et al. 1996, *Kriminalpolizeiliche Auswertung, Bestandsaufnahme und konzeptionelle Überlegungen zur kriminalpolizeilichen Auswertung im Bundeskriminalamt*; zitiert nach AHLF 2002, S. 3.

zur kriminalpolizeilichen Auswertung insgesamt elf einzelne Schritte (siehe Synopse in Anlage B) eines modernen Auswertungsprozesses/Informationsverarbeitungsprozesses zusammen. Dieser Prozess umfasst im Übrigen zunächst die Zieldefinition, die über eine Hypothesengenerierung abgedeckt wird, eine Datenerhebungsphase und eine Datenanalysephase, in der die Ergebnisse entsprechend der Fragestellung generiert werden. Darüber hinaus werden in diesem Prozess die Prozessschritte Informationsordnung und -speicherung genannt.

Die drei vorgestellten Prozesse spiegeln die erforderlichen Schritte einer kriminalistischen Auswertung wider. Nicht berücksichtigt sind jedoch die neuen Bedingungen und Herausforderungen bei der Auswertung von Big Data, die sich durch Masse und Vielfalt der Daten sowie der Notwendigkeit, diese schnell auswerten zu müssen, von der klassischen kriminalistischen Auswertung unterscheidet.

Im nächsten Abschnitt wird deshalb das Augenmerk besonders auf die Auswertungsprozesse gerichtet, die sich ausdrücklich mit der Auswertung von Big Data beschäftigen.

3.4.2 Auswertungszyklus Big Data

In der informationstechnischen Literatur werden verschiedene Prozesse zur Auswertung von Daten, insbesondere Big Data, beschrieben. Vor allem im Zusammenhang mit Data Mining, als eine Methode der Big Data-Auswertung, wird immer wieder der Prozess Cross-Industry Process for Data Mining (CRISP-DM) beschrieben. Zu Beginn des Prozesses stehen hier die Erhebung der zu untersuchenden Geschäftsprozesse und die Ableitung der Anforderungen an die Datenanalyse (Business Understanding). Danach werden Daten gesammelt und im Sinne der Anforderung strukturiert (Data Understanding). Im Anschluss daran erfolgt die *Data Preparation*-Phase. In dieser Phase werden die Daten für die Analyse vorbereitet, die im nachfolgenden Schritt *Modeling* stattfindet. Bei einer Evaluation der Analyseergebnisse (Phase *Evaluation*) werden diese auf Nutzen im unternehmerischen Kontext geprüft. Diese Evaluation kann wiederum Auswirkungen auf das Business Understanding haben und neue Informationsbedürfnisse hervorrufen, womit der CRISP-DM-Zyklus wieder von vorne beginnt. In der letzten Phase, wenn ein erneutes Durchlaufen des Zyklus‘ nicht mehr erforderlich ist, erfolgt die *Deployment*-Phase. In dieser Phase werden die Dokumentation zum Analyseprozess erstellt und die Ergebnisse kommuniziert.[229]

Neben den beschriebenen Prozessen von Walter (Kriminalistik) und CRISP-DM (Informatik) gibt es noch eine Vielzahl weiterer Analysezyklen, wie z. B. der OODA Loop[230] oder der Intelligence Cycle von John Boyd[231], den nachrichtendienstlichen

[229] Vgl. LANQUILLON/MALLOW 2015a, S. 68 ff.
[230] Die Abkürzung OODA steht für Observe, Orient, Decide, Act; HUMMELTENBERG 2010, S. 22.
[231] Vgl. HUMMELTENBERG 2010, S. 22.

Zyklus des schweizerischen Auslandsnachrichtendienstes[232] oder weitere iterative Prozesse im Bereich der Analyse von großen und komplexen Datenmengen, wie der von Sven Hufnagel und Wulf Kollmann[233] vorgestellte Prozess. Diese sollen an dieser Stelle jedoch nicht im Detail vorgestellt werden. Die Synopse in Anlage B fasst die eben beschriebenen Prozesse zusammen und soll als Grundlage für die Entwicklung eines neuen Auswertungskreislaufes dienen.

3.4.3 Entwicklung eines kriminalistischen Big Data-Auswertungszyklus

Wie die vorangestellte Diskussion gezeigt hat, gibt es verschiedene Zyklen, die die Auswertung von Daten zum Ziel haben. Neben den Kernprozessschritten Zielbildung, Datenerhebung und Datenauswertung, die jeder dieser Prozesse aufweist, haben alle Vorgehensweisen eigene Spezifika. Selbst wenn die gesamten Prozessschritte in einem Prozess vereint werden würden, können sie nicht den besonderen Anforderungen einer Big Data-Auswertung im kriminalistischen Umfeld genügen, da die besonderen Rahmenbedingungen des Datenschutzes (siehe Kapitel 3.3.3) und der Datenqualität (siehe Kapitel 3.3.2) nicht berücksichtigt werden würden. Daher wird nachfolgend ein Modell entwickelt und im nächsten Kapitel ausführlich dargestellt, welches diese Belange berücksichtigt. Bei diesem neuen Modell handelt es sich um eine Synthese aus verschiedenen Modellen der Auswertung aus der Kriminalistik und Informatik, welche um weitere Aspekte ergänzt wird, die sich erst durch die Kombination beider Bereiche ergeben.

Der hier zu beschreibende kriminalistische Big Data-Auswertungszyklus (kBDA) besteht aus folgenden acht Phasen:

Definition der Ziele und Planung der Maßnahmen: Im ersten Schritt erfolgt die Definition des Auswertungsziels. Dieser Prozessschritt ist in allen diskutierten Prozessen enthalten und ist ebenfalls für den kBDA unerlässlich, da eine Erhebung von personenbezogenen Daten einen Grundrechtseingriff darstellt und in Grundrechte nur eingegriffen werden darf, wenn der Eingriff erforderlich ist. Diese Prüfung ist in diesem Prozessschritt vorzunehmen. Aufbauend auf der Zieldefinition erfolgt in diesem Schritt auch die Maßnahmenplanung.

Erhebung der Daten und Bewertung der Quellen: Nun erfolgt die operative Datenerhebung durch entsprechende strafprozessuale Eingriffe oder Umwidmung bereits erhobener Daten. Nach der Erhebung erfolgt die Bewertung der Datenquellen.

Data Profiling: Hierbei handelt es sich um die erste spezifische Big Data-Prozess-Maßnahme, die eine Ersteinschätzung der erhobenen Daten zum Ziel hat.

[232] Vgl. CIVELLI 2010c, S. 131.
[233] Vgl. HUFNAGEL/KOLLMANN 2015, S. 147.

Eine solche Maßnahme ist in keiner der beschriebenen Prozesse enthalten. Sie ist jedoch unerlässlich wegen der zu erwartenden heterogenen Datenqualität und -mengen. Aus dieser Einschätzung sollen die weiteren Verarbeitungsprozesse abgeleitet bzw. weitere Datenerhebungen initiiert werden, wenn bereits hier erkennbar ist, dass die vorliegenden Daten nicht als Basis zur Erreichung des angestrebten Ziels genügen.

Aufbereitung der Daten: Der Prozessschritt der Datenverarbeitung ist ein rein informationstechnischer Vorgang, der ebenfalls im CRISP-DM und in einer abgeschwächten Form in den kriminalistischen Auswertungszyklen enthalten ist. Im Rahmen des kBDA umfasst er die Datenextraktion der für die Analyse und Visualisierung erforderlichen Daten, die Transformation der Daten in ein auswertbares Format und das Laden der Daten in eine Analyseanwendung bzw. die Ablage der Daten in einem Zwischenspeicher.

Analyse der Daten: Die Datenanalyse stellt den eigentlichen Kernprozess des kBDA dar. Hier werden durch entsprechende Analysemethoden die Ergebnisse entsprechend der Zielbildung generiert.

Darstellung der Ergebnisse: Die Ergebnisdokumentation wird lediglich von Büchner et al. explizit angeführt (siehe Synopse) und kann in den anderen Prozessen implizit angenommen werden, da eine Ergebnisdarstellung im Rahmen eines Ermittlungsverfahrens erforderlich ist (siehe Kapitel 4.6). Ziel ist die Weitergabe der neu gewonnenen Informationen an andere Verfahrensbeteiligte, um aus ihnen für das Verfahren relevantes Wissen zu gewinnen.

Dokumentation des Prozesses: Dieser Prozessschritt ergibt sich ebenfalls aus der Dokumentationspflicht im Ermittlungsverfahren. Hierdurch soll festgehalten werden, wie das Ergebnis erzielt worden ist und soll alle Verfahrensbeteiligten (z. B. Strafverteidiger) in die Lage versetzen, die für sie notwendigen Informationen (z. B. über Verfahrens- und Beurteilungsfehler) zu gewinnen.

Evaluation des Ergebnisses und des Prozesses: Der letzte Schritt ist die Prozessevaluation, in der das angestrebte Auswertungsziel mit dem erzeugten Analyseergebnis überprüft wird. Hieraus könnte sich das Erfordernis ergeben, entweder den gesamte Prozess oder Teilprozesse nochmals zu durchlaufen.

Die Abbildung 2 stellt diese Schritte im Zusammenhang, Prozessverlauf und Abhängigkeiten dar.

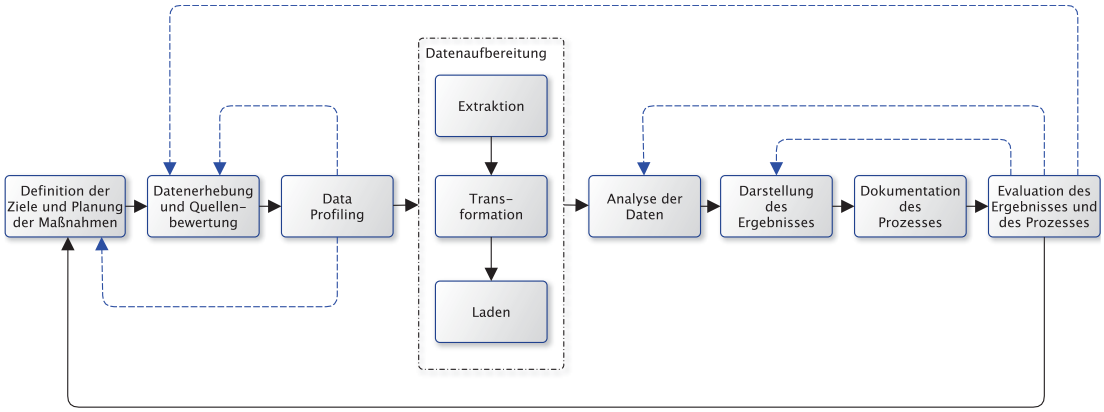

Abbildung 2: Der kriminalistische Big Data-Auswertungszyklus[234]

4 Der kriminalistische Big Data-Auswertungszyklus

Im letzten Kapitel wurde festgestellt, dass die bereits bestehenden Auswertungsprozesse nicht den Ansprüchen einer kriminalistischen Big Data-Auswertung genügen können. Daher wurde ein für dieses Aufgabenfeld spezialisierter Prozess entwickelt. Dieser soll nun an dieser Stelle detailliert erörtert werden. Eine grafische Darstellung des gesamten kBDA – einschließlich der Daten- und Informationsflüsse – ist in Anlage C abgebildet.

4.1 Definition der Ziele und Planung der Maßnahmen

Für Hans Walder und Thomas Hansjakob[235] ist der Auslöser eines Auswertungsprozesses ein Verdacht, der zugleich Ausgangspunkt des kriminalistischen Denkens ist. Dabei ist der Verdacht nicht nur auf den Anfangsverdacht der zugrundeliegenden Straftat begrenzt, sondern umfasst alle Verdachtsmomente innerhalb der Ermittlungen. Die Verdachtslage kann daher als Zustand einer unvollständigen und unbestätigten Informationslage verstanden werden, die im Verlauf einer Auswertung vervollständigt bzw. bestätigt werden soll. Ebenso sehen auch die anderen vorgestellten Auswertungsprozesse die Festlegung der Zielrichtung der Auswertung durch eine Identifizierung fehlender Informationen als ersten Schritt vor (siehe Synopse in Anlage B). Das Durchdenken des Sachverhaltes und die Beschäftigung mit den fehlenden Elementen ist Teil der kriminalistischen Aufgabe.[236] Diese umschließt aber nicht nur die Identifikation der Informationslücken, sondern beinhaltet auch die Festlegung der Methodik zu deren Schließung.

Daher ist der erste Schritt des kriminalistischen Big Data Auswertungsprozesses die Zieldefinition, in der eben diese Informationslücken identifiziert werden sollen.

[234] Abbildung 2: eigene Darstellung.
[235] Vgl. WALDER/HANSJAKOB 2016, S. 93.
[236] Vgl. ACKERMANN/CLAGES/ROLL 2011, S. 166.

Zudem sollen in dieser Phase die geeigneten und erforderlichen Maßnahmen geplant werden, um die benötigten Informationen zu erhalten.

Der gesamte Prozess der Zielbildung und Maßnahmenplanung lässt sich wie in Abbildung 3 darstellen.

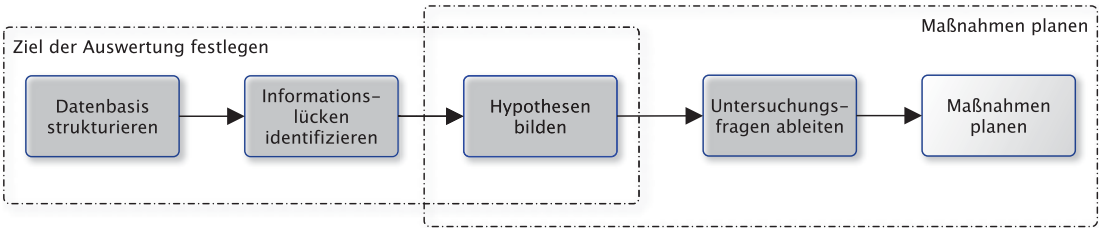

Abbildung 3: Prozess der Zielbildung und Maßnahmenplanung[237]

Ausgangspunkt ist dabei die Strukturierung der bereits vorliegenden Verfahrensdaten. Erst wenn die Daten systematisiert vorliegen, kann eine weitere Datenerhebung geplant werden, um die Sachverhaltsinformationen zu erheben, die nicht oder noch nicht gesichert vorliegen. Rolf Ackermann[238] empfiehlt zur Klärung dieser Frage die kriminalistische Fallanalyse. Dieser Empfehlung kann ohne weiteres gefolgt werden, erlaubt doch die Fallanalyse die umfängliche und systematische Aufarbeitung des Kenntnisstandes zum untersuchten Sachverhalt nach einem fest definierten Schema.

In der Literatur werden verschiedene Fallanalyseraster beschrieben, die jedoch aufgrund des begrenzten Umfangs dieser Arbeit nicht vollständig dargestellt werden können. Exemplarisch soll daher nur die Fallanalyse von Rolf Ackermann besprochen werden.

Die kriminalistische Fallanalyse nach Ackermann[239] umfasst folgende Gliederungspunkte:

1. **Verdachtslage**
2. **Tatsituation**; mit Ausführungen zum Tatort, Tatzeit, Modus Operandi, Tatmittel, Beute, Tatmotiv, Opfer, Taterdächtige(r) und die Verdachtslage gegen bestimmte Personen
3. **Beweislage**; mit Ausführungen zum Personal- und Sachbeweis
4. **Tat- und Täterhypothese**
5. **Fahndungslage**
6. **Rechtslage**

Die Punkte im Bereich 1 (Verdachtslage) und 2 (Tatsituation) beschreiben alle wesentlichen Informationsgruppen, die für die Abarbeitung eines kriminalistischen

[237] Abbildung 3: eigene Darstellung.
[238] Vgl. ACKERMANN 2005a, S. 462.
[239] Vgl. ACKERMANN 2017, S. 139.

Sachverhaltes geklärt werden sollten. Sind zu einzelnen Punkten keine oder unzureichende Informationen vorhanden, so besteht dort eine Informationslücke, die es durch weitere Ermittlungen zu schließen gilt. Die unter Punkt 3 einzusortierenden Erkenntnisse zu den Beweismitteln müssen hierbei einen Bezug zu den Unterpunkten zu Punkt 2 haben. Mit Hilfe dieser Strukturierung können Lücken der Informationslage (Was ist passiert?) und der Beweislage (Wie kann es bewiesen werden?) festgestellt werden.

Nachdem die Informations- und Beweislücken offengelegt sind, müssen diese durch geeignete Maßnahmen geschlossen werden. Die Grundlage für die Planung sind Hypothesen, welche die Informationslücken im Fallanalyseraster mit Annahmen schließen. Diese Hypothesen stellen verschiedene Erklärungsversuche dar, wie sich z. B. die Tat zugetragen haben könnte, was den Täter zu dieser Tat bewog oder wer als Täter in Frage kommen könnte.[240] Dabei müssen die Hypothesen ihren Ankerpunkt in dem vorliegenden Informationsbestand haben.[241] Eine wesentliche Anforderung ist, dass die Auswahl der Hypothesen vielseitig ist, um eine einseitige Planung der nachfolgenden Datenerhebung und Auswertung zu vermeiden. Vielseitige Hypothesen können die Auswirkungen von verschiedenen psychologischen Verzerrungen – wie z. B. die kognitive Dissonanz (vgl. hierzu Kapitel 3.3.1) – reduzieren. Auch hinsichtlich der Verpflichtung für Staatsanwaltschaft und Polizei, auch entlastend zu ermitteln (§ 160 Abs. 2 StPO), sind verschiedene – vor allem auch gegensätzliche – Hypothesen zu bilden.

Die so erstellten Hypothesen bilden die Basis für die Erstellung von Untersuchungsfragen, deren Beantwortung zur Prüfung der Hypothesen dient. Sie präzisieren die erstellten Annahmen und stehen mit ihnen in einem direkten Zusammenhang. Die Untersuchungsfragen sind dabei so zu formulieren, dass sie durch konkrete Ermittlungshandlungen beantwortet werden können.[242] Bei der Formulierung der Untersuchungsfragen ist zu beachten, dass man oft mehr als eine Frage benötigt, um die zugrundeliegende Hypothese zu klären.[243]

Aus den Untersuchungsfragen werden in einem letzten Schritt die notwendigen Maßnahmen abgeleitet, um die Fragen beantworten zu können und damit die Hypothese zu belegen oder zu verwerfen.[244] Im Falle der Datenauswertung ist zu prüfen, welche Datenanalyse die Untersuchungsfrage am wahrscheinlichsten beantworten kann. Das bedeutet, dass nicht nur die rechtlich zulässigen Maßnahmen abgeleitet werden müssen, um Rohdaten zu erheben. Bei der Wahl der Maßnahme muss zusätzlich bereits im Vorfeld beurteilt werden, ob die gewonnenen Daten durch eine zur Verfügung stehende Analysemethode tatsächlich zur Klärung der

[240] Vgl. ACKERMANN 2005b, S. 542.
[241] Vgl. MUSOLFF/HOFFMANN 2006, S. 100.
[242] Vgl. ACKERMANN/RACHOW 1989, S. 62 f.
[243] Vgl. ACKERMANN/STRAUSS 1986, S. 103.
[244] Vgl. ACKERMANN/CLAGES/ROLL 2011, S. 194.

Untersuchungsfragen ausgewertet werden können.

An dieser Stelle wird deutlich, dass zu den eben beschriebenen Fallhypothesen zum Ereignis weitere auf die Daten und Datenanalyse bezogene Hypothesen aufgestellt werden müssen. So ist bei der Maßnahmenplanung auch auf Aspekte von Datenqualität und Datenmenge einzugehen, damit nicht Daten erhoben werden, deren Erhebung zwar grundsätzlich zulässig und zur Beantwortung theoretisch geeignet wären, aber wegen technischer Gründe oder begrenzter personeller oder auch technischer Ressourcen letztlich nicht ausgewertet werden können. Diese Anforderung setzt ein entsprechendes Fachwissen bei der Verfahrensführung voraus, welches auch über Beratung von Experten beigesteuert werden kann.

Diese Experten müssen in der Lage sein, kriminalistisches Fachwissen, rechtliche Hintergründe sowie Kenntnisse aus der Informatik zu vereinen. Eine Analyse der Curricula der Fachhochschulen von Bund und Ländern, die für die Ausbildung der Polizeien zuständig sind, hat gezeigt, dass dieses Wissen im Vorbereitungsdienst nicht vermittelt wird. Einzelne Bundesländer, wie z. B. Hamburg vermitteln zumindest Basiswissen im Bereich der Informatik.[245] Dieses kann jedoch nicht den Anforderungen dieser komplexen Thematik genügen.

4.2 Erhebung der Daten und Bewertung der Quellen

Auf Grundlage von Zieldefinition und Maßnahmenplanung werden die für die Beantwortung der Untersuchungsfragen benötigten Daten erhoben. Dazu kommen verschiedene Eingriffsmaßnahmen in Betracht, die bereits beispielhaft unter Punkt 3.2 dargestellt wurden. Da der Schwerpunkt dieser Arbeit die Auswertung der erhobenen Daten ist, werden die Maßnahmen an dieser Stelle nicht im Detail beschrieben.

Hans Walder[246] führt in seinem kriminalistischen Zyklus zum Punkt *Daten beschaffen* aus, dass die kriminalistische Arbeit abgeschlossen sei, wenn es gelänge, die fehlenden Daten vollständig zu erheben. Dabei ließ er offen, an welcher Stelle diese Prüfung stattfinden soll. Zumindest im Umfeld von Big Data ist eine solche Prüfung in dieser Phase der Datenerhebung noch nicht möglich, da erst nach einer Analyse der Daten erkennbar ist, ob die Daten den im ersten Schritt aufgestellten Auswertungszielen genügen.

Der Prozessschritt Datenerhebung des kBDA ist mit dem Eingang der Daten noch nicht abgeschlossen. Um später nachvollziehen zu können, wie belastbar und zuverlässig die gewonnenen Daten und die daraus generierten Auswertungsergebnisse sind, muss die Zuverlässigkeit der Datenquellen und deren Zugang zu den konkreten Daten nach einem standardisierten Raster bewertet werden. Die

[245] Vgl. AKADEMIE DER POLIZEI HAMBURG 2017, S. 143.
[246] Vgl. WALDER/HANSJAKOB 2016, S. 94.

Notwendigkeit ergibt sich aus den Kriterien Variety in Verbindung mit Veracity wonach die Datenquellen im Umfeld von Big Data unterschiedliche fachliche Qualität aufweisen können. Zwar werden die auszuwertenden Daten z. B. aus Datenbanken gewonnen, die Befüllung dieser kann jedoch durch unterschiedliche Nutzer erfolgen. So müssen aus fachlich-inhaltlicher Sicht Datenbankauszüge von Telekommunikationsprovidern oder öffentlichen Stellen – wie z. B. dem Kraftfahrtbundesamt – anders bewertet werden als eine Datenbank in einem sozialen Netzwerk, die durch eine unbekannte öffentliche Community befüllt wurde.

Die Einordnung sollte dabei möglichst nah an der Quelle erfolgen[247] und muss mit den Informationen mitgeführt werden, um so „[...] eine schrittweise ‚Verwahrheitung' von Einzelinformationen durch das einsetzende ‚Stille-Post-Prinzip' [...] auf diese Weise weitestgehend [zu unterbinden] [...]."[248]

Für die Beurteilung der Belastbarkeit einer Information kann die sogenannte 4x4 Methode genutzt werden, um mit möglichst objektiven Kriterien zu messen, wie sicher eine Information ist.[249] Diese Metainformation dient zum einen der Weitergabe an andere Nutzer und zum anderen der eigenen Objektivierung. Dazu wird eine Quelle im Zusammenhang mit der gelieferten Information in zwei Dimensionen bewertet.[250] Die erste Dimension bewertet retrograd die Zuverlässigkeit der Informationsquelle und vergibt hierfür eine der vier Stufen A, B, C, oder X (siehe Tabelle 2).

Tabelle 2: Zuverlässigkeit der Quelle[251]

Kategorie	Beschreibung
A	Es bestehen keine Zweifel an der Authentizität, Verlässlichkeit und Eignung der Quelle, oder die Informationen stammen von einer Quelle, die sich in allen Fällen als zuverlässig erwiesen hat.
B	Quelle, deren Informationen sich in den meisten Fällen als verlässlich erwiesen haben.
C	Quelle, deren gelieferte Informationen sich in den meisten Fällen als nicht verlässlich erwiesen haben.
X	Die Verlässlichkeit der Quelle kann nicht eingestuft werden.

Die zweite Dimension bewertet den Abstand der Quelle zu der von ihr übermittelten Information (siehe Tabelle 3). Die Kombination aus beiden Werten ergibt einen Kodierungswert, der angibt, wie sicher die bewerteten Informationen sind. So nimmt man an, dass es sich bei den Kombinationen A1, A2, B1sowie B2 um

[247] Vgl. SPANG 2008, S. 71.
[248] NONNINGER 2002, S. 7.
[249] Vgl. NONNINGER 2002, S. 7.
[250] Vgl. ROLL 2013, S. 363.
[251] ROLL 2013, S. 363.

Tabelle 3: Abstand von Quelle zur Information[252]

Kategorie	Beschreibung
1	Eigene Wahrnehmung oder sonstige zweifelsfreie Herkunft
2	Andere als polizeiliche Quelle, die direkten Zugang zur Information hatten
3	Information von Hörensagen, die sich mit anderen Informationen deckt
4	Unbestätigte Information von Hörensagen

sichere Informationen handelt, und die übrigen Informationen als ungesichert gelten.[253]

Da Ermittlungsmaßnahmen in Grundrechte eingreifen können, z. B. Telefonüberwachungen, Durchsuchungen und Erhebungen von Verkehrsdaten, muss der Belastbarkeit von neuen Daten ein besonderer Wert zugerechnet werden. Damit erfährt die hier beschriebene Einstufung einen hinweisenden, teilweise grundrechtsschützenden Charakter, indem diese aufzeigt, dass die neuen Erkenntnisse auf einer unsicheren Datenbasis gründen. Diese Unsicherheit kann durch die Verarbeitung von unsicheren Daten über mehrere Stufen ‚vererbt' werden (siehe Abbildung 4). In diesen Fällen müssen die Erkenntnisse durch weitere Maßnahmen abgesichert werden.

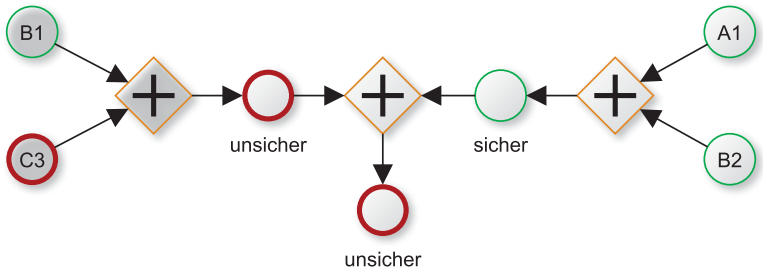

Abbildung 4: Vererbung von Quellenbewertungen bei Datenzusammenführung[254]

4.3 Data Profiling

Nach der Erhebung und der Bewertung von Informationen erfolgt der erste Schritt, welcher der Datenverarbeitung im engeren Sinne[255] zugeordnet werden kann. Die-

[252] ROLL 2013, S. 363.
[253] Vgl. AHLF 2002, S. 5.
[254] Abbildung 4: eigene Darstellung.
[255] „Als Datenverarbeitung (im engeren Sinne) werden Vorgänge bezeichnet, die eine manuelle oder maschinelle (automatisierte) Sortierung, Selektierung und/oder Umformung von Eingabedaten in Ausgabedaten zum Inhalt haben und die als
- Stapelverarbeitung und/oder
- Dialogverarbeitung

ser wird Data Profiling genannt und dient dazu, die Datenqualität dahingehend zu prüfen, ob die gewonnenen Ausgangsdaten für eine Analyse geeignet sind. Dabei werden vorrangig Metadaten über den vorliegenden Datenbestand gewonnen bzw. bestehende Metadaten korrigiert. Es geht hier explizit nicht um eine fachliche Auswertung oder eine qualitative Veränderung der Daten.[256] Data Profiling soll den Analysten und Ermittlern ein objektives Bild über ihre Daten verschaffen, um entsprechende Maßnahmen wie z. B. Datenneubeschaffungen oder Qualitätsverbesserungen ableiten zu können.

Dieses Erfordernis ergibt sich aus den Eigenschaften Variety sowie Veracity von Big Data und den daraus entstehenden Risiken für die Auswertung aufgrund der Datenqualität. Apel et al. beschreiben Data Profiling als „[...] weitgehend automatisierte[n][...] Prozess zur Analyse vorhandener Datenbestände. Verschiedene Analysetechniken liefern Informationen über Inhalt, Strukturen und Qualität der Daten.“[257] Solche automatisierten Prozesse sind aber nur dann möglich, wenn es sich um gleichbleibende Datenquellen mit gleich angelieferten Datenstrukturen handelt. Diese sind im wirtschaftlichen Umfeld (z. B. hauseigene Buchungssysteme) aber auch im kriminalistischen Kontext (z. B. polizeieigene Datenbanken) vorzufinden. Sobald Datenstrukturen erstmalig bearbeitet werden oder sich diese verändern, ist die Nutzung vordefinierter vollautomatisierter Prozesse nicht mehr möglich. Neben der Einschätzung der Datenqualität dient Data Profiling auch der Erkenntnis, wie die Quelldaten zusammenhängen.[258] Das ist vor allem dann vonnöten, wenn die fachlich zusammenhängenden Daten technisch über mehrere Speicherorte verteilt sind. Erst durch Aufhellung der Datenspeicherungsstrukturen ist in diesen Fällen eine zielgerichtete Analyse der Daten überhaupt erst möglich.

Data Profiling ist ein mehrstufiges iteratives Verfahren, bestehend aus den Schritten: Daten integrieren, integrierte Daten analysieren, Ergebnisse darstellen und Ergebnisse fachlich bewerten.[259]

In Schritt 1 werden zunächst alle erforderlichen Daten extrahiert und aufbereitet (**Datenintegration**). So können z. B. zusammenhängende Ortsangaben in einzelne Detailinformationen (Postleitzahl, Ort und Straße) aufgespalten werden.[260] Diese Vorbereitungstechniken haben starke Parallelen zu dem im Anschluss beschriebenen Datenaufbereitungsprozess. Daher können Teile der dabei generierten Datenbestände und Techniken (z. B. Programmiercode) im anschließenden Aufbereitungsprozess wiederverwendet werden.

Nach Export und Aufbereitung der Daten erfolgt in Schritt 2 die **Analyse der Da-**

gestaltet sein kann.“ KÄNEL o. D. S. 3.
[256] Vgl. APEL/BEHME et al. 2015, S. 131.
[257] APEL/BEHME et al. 2015, S. 131.
[258] Vgl. SCHNIDER et al. 2016, S. 64.
[259] Vgl. APEL/BEHME et al. 2015, S. 132.
[260] Vgl. APEL/BEHME 2010, S. 117.

ten zur Datenqualitätsprüfung und zum Erkennen von Abhängigkeiten und Zusammenhängen der Daten untereinander. Aufgrund der Komplexität einer solchen Analyse sollten hierfür geeignete Werkzeuge eingesetzt werden.[261] Im polizeilichen Umfeld werden u. a. die kommerzielle Anwendung HumanIT InfoZoom[262] oder die OpenSource-Anwendung KNIME[263] genutzt. Data Profiling bedient sich dabei verschiedener Einzelanalysen.[264] Die nachfolgend kurz dargestellten Data Profiling-Analysen können auch im Rahmen der kriminalistischen Big Data-Auswertung zum Tragen kommen.

Attributnamenanalyse: Bei dieser Analyse werden die Datenfeldnamen von strukturiert vorliegenden Datenbeständen auf Verständlichkeit und Eindeutigkeit überprüft.[265] Insbesondere bei Daten aus verschiedenen Quellen, die später zusammen fachlich ausgewertet werden sollen, kann hier festgestellt werden, ob entsprechende Daten aus den verschiedenen Quellen vorliegen.

NULL-Werte-Analyse: Mit Hilfe der NULL-Werte-Analyse wird die Vollständigkeit der Datenanlieferung geprüft. Hierbei werden Datenfelder nicht nur auf nicht vorhandene Daten (NULL-Wert) geprüft, sondern auch auf Ersatzwerte, die nicht den erwarteten Werten (z. B. ein Leerzeichen oder Bindestrich im Datenfeld für Rufnummern) entsprechen.[266]

Wertebereichsanalyse: Im Rahmen der Wertebereichsanalyse wird die Verteilung der einzelnen Werte eines einzelnen Datenfeldes ermittelt.[267] Eine besondere Aufmerksamkeit erfahren hier die Datenfelder für Zeitangaben. So müssen z. B. die Datensätze einer Funkzellenauswertung auf den angeforderten Zeitraum und den korrekten Standort der Funkzelle überprüft werden. Zudem wird bei der Wertebereichsanalyse auch die Genauigkeit der Werte überprüft. Hierunter fällt auch die Prüfung, ob die Werte aus verschiedenen Quellen den gleichen Werteskalen, wie z. B. gleiche Währung, gleiche Längeneinheit oder gleiche Zeitzonen, zugeordnet sind.

Im dem auf die Analyse folgenden Schritt sind die Ergebnisse der Data Profiling-Analyse in geeigneter Form und für die anderen Verfahrensbeteiligten verständlich zu dokumentieren (**Ergebnisdarstellung**). Diese Dokumentation fließt in die Prozessdokumentation des kriminalistischen Big Data-Auswertungsprozesses ein (siehe Kapitel 4.7).

[261] Vgl. APEL/BEHME et al. 2015, S. 174; vgl. SCHNIDER et al. 2016, S. 64.
[262] Vgl. HABERBERGER/TALARCZYK 2007, S. 368.
[263] Vgl. STOCK 2017, S. 20 f.; Homepage von KNIME: https://www.knime.com (besucht am 01.01.2018).
[264] Vgl. APEL/BEHME et al. 2015, S. 137 ff.
[265] Vgl. APEL/BEHME et al. 2015, S. 137.
[266] Vgl. APEL/BEHME et al. 2015, S. 149.
[267] Vgl. APEL/BEHME et al. 2015, S. 142.

Zum Abschluss des Data Profiling-Prozesses erfolgt die **fachliche Bewertung** der Ergebnisse und Klärung möglicher offener Fragen, um schließlich eine Bewertung im Hinblick auf die Verwendbarkeit der Datenquellen für die anschließende Analyse zu treffen. Apel und Behme[268] schlagen vor, diese Bewertung mit den entsprechenden fachlichen Experten – in Strafverfahren sind das die polizeilichen Ermittler und die Staatsanwaltschaft – durchzuführen. Die Notwendigkeit einer solchen Beratung ist im Rahmen eines Ermittlungsverfahrens abhängig vom Grad der Arbeitsteilung und der Spezialisierung der einzelnen Mitglieder des Ermittlungsteams.

Sollte bereits beim Data Profiling festgestellt werden, dass eine Bereinigung bzw. Verbesserung der Datenqualität im weiteren Datenverarbeitungsprozess nicht möglich ist und die Daten für eine Analyse ungeeignet sind, wären eventuell neue Daten zu erheben oder in schwerwiegenden Fällen der Auswertungsprozess an dieser Stelle abzubrechen. Daher nimmt der Prozessschritt *Data Profiling* eine entscheidende Position im kriminalistischen Big Data-Prozess – insbesondere in der Verfahrens- und Ressourcenplanung – ein.

4.4 Aufbereitung der Daten

Der informationstechnische Kernprozess im Rahmen der Datenauswertung ist der Datenaufbereitungsprozess – auch ETL-Prozess[269] genannt – mit all seinen Unterprozessen; Extraktion aus den Quellsystem, Transformation der Datenstrukturen und Laden der Daten in ein Zielsystem. Diese Arbeitsschritte beanspruchen einen erheblichen Teil der Bearbeitungszeit. Lanquillon und Mallow[270] führen dazu aus, dass im Zusammenhang mit Datenaufbereitung von Big Data mindesten 70-80 % der Arbeitszeit hierauf entfallen. Das Ziel des ETL-Prozesses ist die Vereinheitlichung und Herstellung von Konsistenz von Daten aus den unterschiedlichen Datenquellen. Dabei werden die Daten dauerhaft für den entsprechenden Zweck geändert.[271] Durch eine Standardisierung der Datenfelder und -strukturen sollen „[. . .] semantisch identische Inhalte identisch dargestellt werden.“[272] Das Ergebnis ist ein vergleichbarer Datenbestand, der ursprünglich in verschiedenen Datenquellen mit unterschiedlichen Abbildungsregularien gespeichert war.

Der Umfang und die Richtung des ETL-Prozesses richten sich dabei an zwei Einflussgrößen aus. Die erste ist das Auswertungsziel, welches im Schritt der Zieldefinition im Rahmen der Hypothesenbildung und Ermittlungsfragenerstellung festgelegt wurde. Diese Einflussgröße bedingt, dass nur die für die Auswertung rele-

[268] Vgl. APEL/BEHME 2010, S. 117.
[269] Die Abkürzung ETL steht für *Extraktion, Transformation, Laden*.
[270] Vgl. LANQUILLON/MALLOW 2015a, S. 79.
[271] Vgl. MERTENS/WIECZORREK 2000, S. 145.
[272] BLEIHOLDER/SCHMID 2015, S. 138.

vanten Daten im ETL-Prozess verarbeitet werden. Andere Daten werden vom weiteren Prozess ausgeschlossen, soweit sie nicht aus anderen Gründen, wie z. B. Datenschutz-, Dokumentations- oder Visualisierungszwecken, als Metadaten im Prozess mitgeführt werden müssen. Die zweite Einflussgröße ist die Analysesanwendung, mit deren Hilfe die Daten analysiert werden sollen.[273] Diese bestimmt den Datenumfang (Datentiefe und Datenbreite) und das Datenformat, welches im ETL-Prozess zu generieren ist. Da sich diese beiden Anforderungen bei jedem einzelnen Fall ändern können, handelt es sich bei dem ETL-Prozess um einen dynamischen Prozess.

Nachfolgend wird der ETL-Prozess im Einzelnen vorgestellt, da dieser den Kernprozess in der Datenvorbereitung kriminalistischer Big Data-Bestände darstellt.

4.4.1 Extraktion der Daten

Die Datenextraktion dient der Selektion der zu analysierenden Daten aus verschiedenen Quellen.[274] Aufgrund dieser Verschiedenartigkeit der Quellen kann die Datenextraktion über unterschiedliche Wege erfolgen. Einerseits können Daten per Datenexport über eigens für den Export vorgesehenen Exportschnittstellen oder durch Export z. B. als Text- oder Exceldateien erfolgen.[275] Andererseits können aber auch Kopien ganzer Datenbanken, sogenannte Database Dump, als Quelle im ETL-Prozess dienen.[276] Während in der Wirtschaft bei der Extraktion von Daten aus laufenden Datenbankanwendungen die Verwendung bestimmter Technologien erforderlich sind, weil diese gleichzeitig im operativen Betrieb benötigt werden,[277] erfolgt die Datenextraktion in Strafverfahren überwiegend aus statischen Daten, weil diese Daten zu einem bestimmten Zeitpunkt sichergestellt bzw. beschlagnahmt wurden und eine Veränderung nicht mehr erfolgt. Dennoch können auch in einem Ermittlungsverfahren dynamische Datenquellen – wie z. B. laufende TKÜ-Anwendungen – herangezogen werden, die ein vergleichbares Vorgehen wie in der Wirtschaft erfordern. In diesen Fällen muss der Umfang und vor allem die Frequenz der Datenextraktion festgelegt werden. So besteht die Möglichkeit, bedarfsorientiert oder periodisch den gesamten Datenbestand (Full Extraction) oder nur eine Teilmenge der Daten zu exportieren, die im Zeitraum zwischen dem letzten Export und der aktuellen Extraktion verändert wurden (Delta Extraction).[278] Die Wahl, ob der gesamte oder nur ein definierter Teil des Datenbestandes exportiert wird, ist vom Auswertungsziel, vom Datenumfang und der Exportfunktionalität des Quellsystems abhängig.

[273] Vgl. HABERBERGER/TALARCZYK 2007, S. 368.
[274] Vgl. FARKISCH 2011, S. 60.
[275] Vgl. HABERBERGER/TALARCZYK 2007, S. 368.
[276] Vgl. FARKISCH 2011, S. 62.
[277] Vgl. FARKISCH 2011, S. 62.
[278] Vgl. SCHNIDER et al. 2016, S. 67.

Als Datenquellen kommen aber nicht nur Datenbanken in Betracht. Neben diesen können andere Datenquellen für eine Auswertung von Interesse sein. Vor allem bei Internetinhalten, aber auch bei anderen Quellen mit strukturierten, semistrukturierten und auch unstrukturierten Daten kann ein Export von Daten erforderlich sein. Verfügen Anwendungen über keine entsprechende Schnittstelle, so müssen die Daten über andere Technologien wie z. B. Web-Crawler exportiert werden.[279] So können z. B. Internetinhalte über eigens für die Datenextraktion programmierte Bots, wie sie auch von großen Suchmaschinen eingesetzt werden, extrahiert werden.[280]

Zudem fällt – wie in der Wirtschaft auch – in Strafverfahren ein überwiegender Teil der Informationen in unstrukturierten Datenformaten an. Das können z. B. Texte, Bilder, Videos oder Tonspuren sein. Um diese Daten der Auswertung zugänglich zu machen, müssen die darin enthaltenden Daten extrahiert und strukturiert abgelegt werden. Vor allem Kerninformationen wie z. B. Personen-, Orts- und Organisationsnamen, Kfz-Kennzeichen, Bankdaten und Währungsbeträge müssen im kriminalistischen Big Data-Szenario automatisiert extrahiert werden können.[281] Das betrifft nicht nur Daten von sichergestellten Datenträgern, sondern auch Informationen aus den Social Media-Plattformen im Internet.[282] Die Methoden der Extraktion reichen vom einfachen Mustersuchen (siehe Anlage A) bis hin zu hochentwickelten Parsing-Technologien, die Entitäten über komplexe grammatische Strukturen in den strukturlosen Quellen finden können.[283]

4.4.2 Transformation der Daten

Nach der Extraktion der Daten aus dem Quellsystem erfolgt die Datentransformation. Ziel der Transformation ist das Anpassen von Datenstrukturen an die Erfordernisse der Analysesoftware. Dabei erfolgt die Transformation in vier Schritten: Datenvalidierung, Datenstandardisierung, Datenbereinigung und Datenanreicherung.[284]

Voraussetzung für die Datenverarbeitung ist eine **Datenvalidierung**, um potenziell inkonsistente und fehlerhafte Daten zu identifizieren.[285] Zudem erfolgt eine Prüfung der Quelldaten nach vorab definierten Regeln, wie z. B. Datentypen oder Wertebereiche. Entsprechen die Daten den erwarteten Parametern nicht, werden diese entweder verworfen oder müssen aufbereitet werden.[286] Die Techniken zur

[279] Vgl. LANQUILLON/MALLOW 2015b, S. 271.
[280] Vgl. MERTENS/BODENDORF et al. 2017, S. 55.
[281] Vgl. ZIERCKE 2014, S. 15.
[282] Vgl. HELD 2003, S. 3; Vgl. SPRANGER/LABUDDE 2017, S. 167.
[283] Vgl. BITCOM 2014, S. 60.
[284] Vgl. APEL/BEHME et al. 2015, S. 120.
[285] Vgl. FARKISCH 2011, S. 63.
[286] Vgl. APEL/BEHME et al. 2015, S. 120.

Datenprüfung wurden bereits im Schritt des *Data Profiling* eingesetzt, um die Datenqualität des gesamten Datenbestandes auf Geeignetheit für eine Auswertung zu untersuchen. Die Datenvalidierung in der Phase des ETL-Prozesses hingegen dient der Einschätzung der einzelnen Datensätze und der Ableitung von notwendigen Datenstandardisierungsmaßnahmen jeden einzelnen Datensatz.

Die **Datenstandardisierung** betrifft die Struktur der konkreten Daten. Diese umfasst die Abbildung der Datenobjekte wie bspw. einer Anschrift, die in einer Quelle (A) aus (1) Postleitzahl, (2) Ort und (3) Straße und (4) Hausnummer in je einem Datenfeld besteht. In einer anderen Quelle (B) besteht die Anschrift aus (1) Postleitzahl und Ort, sowie (2) Straße und Hausnummer in jeweils einem Datenfeld. Die Analyseanwendung verlangt jedoch zur Abbildung einer Anschrift die (1) Postleitzahl, den (2) Ort und die (3) Straße mit Hausnummer in jeweils einem separaten Datenfeld. In diesem Beispiel müssen die Daten zum einen zusammengeführt (A) und zum anderen aufgetrennt (B) werden (siehe Abbildung 5).

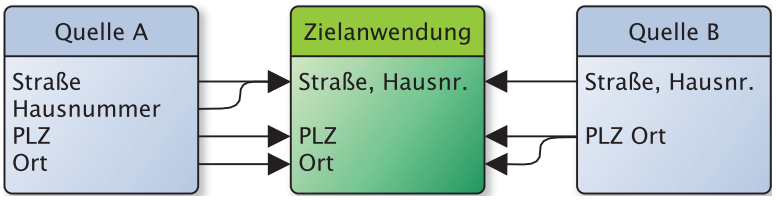

Abbildung 5: Beispiel für eine Datenfeldstandardisierung[287]

Des Weiteren kann sich die Datenstandardisierung auch auf die Datenwerte selbst beziehen. So sind eventuell Katalogwerte anzupassen[288] oder regionale Datendarstellungen (z. B. amerikanisches Datum $\mapsto$ deutsches Datum) anzugleichen[289].

Die Datenstandardisierung befasst sich auch mit der Auflösung komplexer Datenstrukturen und Abbildung der Daten in einem anderen Format. So können in diesem Schritt z. B. HTML-Dateien in ein Tabellenformat oder eine Tabelle in ein XML-Format umformatiert werden. Dabei ist das Zielformat von der Analyseanwendung abhängig. Neben den fachlichen Standardisierungen kann es aus technischer Sicht auch notwendig sein, Anpassungen an den Datenbestand vorzunehmen. Eine solche technische Standardisierung ist die Vereinheitlichung der Zeichenkodierungen (z. B. UTF8, UTF16, ANSI). Diese können dann nötig werden, wenn Daten auf verschiedenen Betriebsplattformen (z. B. MacOS, Linux oder MS Windows) erzeugt wurden.[290]

Neben der Datenstandardisierung müssen gelegentlich auch Daten bereinigt werden. Diese **Datenbereinigung** kann technische und fachliche Gründe haben.

[287] Abbildung 5: Eigene Darstellung.
[288] Vgl. ABTS/MÜLDER 2017, S. 278.
[289] Vgl. FARKISCH 2011, S. 64.
[290] Vgl. APEL/BEHME et al. 2015, S. 76.

Die Bereinigung bezieht sich dabei auf ganze Datensätze und -felder, aber auch auf Dateninhalte. Ein wichtiger Aspekt ist die im Rahmen der Datenqualität angesprochene Redundanzfreiheit. Hiernach darf jeweils nur ein Datensatz zur gleichen Entität in der Zieldatenbank existieren. Durch die Vielfalt der Datenquellen im Big Data-Kontext werden aber Datensätze zur gleichen Entität (z. B. eine Person) mehrfach angeliefert. Dabei handelt es sich nicht unbedingt nur um einen unerwünschten Nebeneffekt. Die redundante Datenzulieferung ist in verschiedenen Auswertungsszenarien gewollt. Bei jeder Anlieferung werden nämlich andere Teilaspekte zum jeweiligen Betrachtungsobjekt mitgeliefert, die schließlich im Rahmen einer Datenfusion in einer Gesamtschau zusammengefügt werden sollen.[291]

Ungeachtet der Gründe der wissentlichen oder unwissentlichen redundanten Datenanlieferung müssen diese Daten jedenfalls zusammengeführt werden. Soweit die eindeutigen Identifizierungsmerkmale (z. B. Kundennummern) einzigartig sind, ist das grundsätzlich unproblematisch. Probleme bei der Datenfusion entstehen jedoch, wenn die Identitätsklärung nicht auf eindeutige Werte zurückgreifen kann. Das wäre z. B. der Fall, wenn das Identifizierungsmerkmal bei Personen der Name ist, da es oft mehrere Personen mit gleichem Namen – auch in der Kombination mit dem Vornamen – gibt.[292]

Aber auch wenn das Identifizierungsmerkmal eindeutig ist, kann es zu Problemen bei der Datenfusion kommen. Diese treten dann auf, wenn gleiche Attribute unterschiedliche Werte zum gleichen Realobjekt (z. B. unterschiedliche Anschriften oder Rufnummern zur gleichen Person) enthalten. Für die weitere Bearbeitung müssen entweder die Objekte verschmolzen werden, wobei unter Umständen dann nicht alle Attributswerte übernommen werden können. Andererseits können auch die redundanten Objekte beibehalten werden, um so auch alle Attributswerte für die Analyse zugänglich zu machen.

Jede mögliche Reaktion auf Dubletten kann weitreichende Auswirkungen auf die Datenanalyse haben. Werden diese Doppelbestände belassen, könnten z. B. fachlich nicht vorliegende Häufungen suggeriert werden (z. B. bei Auswertung der Kontakthäufigkeit). Werden Doppelbestände gelöscht, könnten eventuelle Zusammenhänge nicht mehr sichtbar sein. Im Gegensatz dazu können bei einer ungeprüften automatisierten Datenverschmelzung wiederum falsche Zusammenhänge erzeugt werden. Daher ist der Umgang mit Dubletten vom jeweiligen Einzelfall abhängig.

Alle Veränderungen an den Daten sind zu dokumentieren.[293] Insbesondere in einem Ermittlungsverfahren ist eine Dokumentation von Veränderungen unumgänglich, um die Nachvollziehbarkeit der digitalen Spuren zu gewährleisten.

Apel führt als weiteren optionalen Punkt der Datentransformation die **Datenan-**

[291] Vgl. S. MÜLLER 2016, S. 157.
[292] Vgl. GARFINKEL 2009, S. 94.
[293] Vgl. WEICHERT o.D., S. 15; Vgl. FARKISCH 2011, S. 64.

reicherung an. Hierbei wird der bestehende Datenbestand durch Hinzufügen weiterer Daten veredelt.[294] Diese Datenveredelung erfolgt auf gleiche Weise wie die Datenfusion, wobei es sich hierbei technisch und fachlich um Verschmelzung einer wissentlich herbeigeführten Datenredundanz handelt.

4.4.3 Laden der Daten

Im letzten Schritt des ETL-Prozesses werden die extrahierten und transformierten Daten in die Zielanwendung geladen. Dieser Import in die Zielanwendung kann dabei über mehrere Wege, ausgehend vom ETL-Prozess, erfolgen (siehe Abbildung 6).

Der erste Importweg sieht einen direkten Import der Daten in die Analyseanwendung vor. Dazu wird am Ende des ETL-Prozesses eine Importdatei entsprechend den spezifischen Anforderungen der Analyseanwendung erstellt. Der zweite Im-

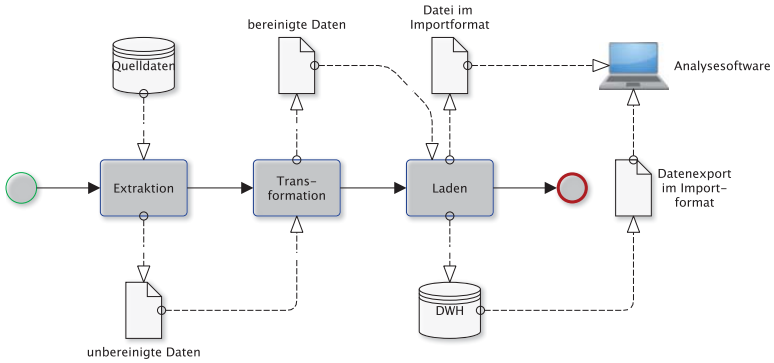

Abbildung 6: ETL-Prozess[295]

portweg führt über eine zentrale Datenbank, in der die Daten für eine Auswertung vorgehalten werden. In der Wirtschaft ist das z. B. ein Data-Warehouse. Das Data Warehouse-Konzept wurde von William H. Inmon erstellt. Er beschreibt das Data Warehouse wie folgt: „A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions.“[296]

Insbesondere die Merkmale Themenorientierung und Zeitvarianz spielen neben der Datenintegration eine wichtige Rolle.

Die **Themenorientierung** (subject-oriented) beschreibt, dass ein Data Warehouse nicht auf einzelne Transaktionen ausgerichtet ist, sondern auf Kennzahlen wie Kosten und Umsätze.[297] Im kriminalistischen Kontext kann hierunter die Sammlung von Ermittlungsdaten für spezifische Auswertungen verstanden werden, wie z. B. für Häufigkeitsanalysen von Verbindungsdaten. Somit wird ein Data

[294] Vgl. APEL/BEHME et al. 2015, S. 121.
[295] Abbilung 6: eigene Darstellung.
[296] INMON 2005, S. 29.
[297] Vgl. FARKISCH 2011, S. 5 f.

Warehouse nicht zur Durchführung des operativen Geschäfts eingesetzt, sondern soll die Sicht auf einen spezifischen – zum Teil übergeordneten – Themenbereich ermöglichen.

Die Eigenschaft **Zeitvarianz** (time-variant) bezieht sich auf die Aktualität der Daten. So sind Daten aus einem operativen IT-System wie z. B. einem Buchungssystem, einem kriminalistischen Fallbearbeitungssystem[298] oder aus Spezialanwendungen wie z. B. TKÜ-Anwendungen aktuell. Im Gegensatz dazu stellen Daten in einem Data Warehouse eine Momentaufnahme der Datenbestände dar, zu der diese ins Data Warehouse aufgenommen wurden. Die Aktualisierung erfolgt in der Regel periodisch. Zur Auswertung von zeitlichen Trends werden hierzu alle periodisch erhobenen Daten über einen längeren Zeitraum aufgenommen und für eine Analyse vorgehalten. Dabei werden die einmal eingebrachten Daten weder modifiziert, noch entfernt, sodass hierdurch eine Historisierung der Daten – also die Möglichkeit, die Veränderung der einzelnen Datensätze nachvollziehen zu können – aus den Quellsystemen erfolgt.[299] So lässt sich der Erkenntnisstand der Ermittlungen – die vollständige Befüllung des Data Warehouse vorausgesetzt – zu verschiedenen Zeitpunkten reproduzieren.

Der wesentliche Unterschied zwischen einem Data Warehouse für Strafverfahrensdaten zu einem Data Warehouse in der Wirtschaft ist der Umstand, dass ein großer Teil der Ermittlungsdaten einmalig an die Behörden geliefert wird. Eine periodische Übermittlung kann überwiegend nur aus den polizeieigenen Datenbeständen erfolgen. Zudem müssen in einzelnen Fällen Daten aus datenschutzrechtlichen Gründen gelöscht oder verändert werden. Somit reduziert sich der Mehrwert eines Data Warehouse im kriminalistischen Umfeld auf die zentrale Datenspeicherung der Informationen in einem einheitlichen Format.

Im Hinblick darauf, dass sich die Datenqualität auch nach dem Datennutzer bemisst (siehe Kapitel 3.3.2) und nicht alle zukünftigen Auswertungsnotwendigkeiten im Vorfeld bedacht werden können, kann es trotz umfangreich durchgeführter Transformationsprozesse erforderlich werden, den ETL-Prozess von bereits verarbeiteten Daten nochmals zu durchlaufen. So kann beispielsweise die Möglichkeit bestehen, dass Attribute einer Entität (z. B. die Farbe eines Kfz), die bisher für eine Auswertung nicht erforderlich waren und deshalb noch nicht in das Data Warehouse aufgenommen wurden, für die Auswertung relevant werden. In diesem Beispiel müssten alle Daten (z. B. die Kfz-Daten) nochmals aufbereitet werden, um das neue Attribute (hier die Farbe des Kfz) ins Data Warehouse aufzunehmen. Hierdurch wird deutlich, dass die Einrichtung und Befüllung eines solchen Data Warehouse gewissenhaft vorbereitet und durchgeführt werden muss.

Liegen letztendlich die Daten entsprechend der Spezifikationen der Analysean-

[298] Vgl. EDER 2004, S. 4 ff.
[299] Vgl. FARKISCH 2011, S. 6.

wendung im Data Warehouse vor, können diese aus dem Data Warehouse exportiert und in die Anwendung zur weiteren Verarbeitung und Analyse geladen werden.

4.5 Analyse der Daten

Nachdem eine vereinheitlichte und bereinigte Datenbasis erstellt wurde, werden die Daten entsprechend der kriminalistischen Fragestellung analysiert.

Dieser Schritt ist der eigentliche fachliche Kernprozess, um aus Daten und Informationen neues Wissen zu generieren. Ziel ist die Beantwortung der im ersten Schritt aufgestellten Untersuchungsfragen, die wiederum die erstellten Hypothesen bestätigen bzw. verwerfen sollen. Ziel der Auswertung ist entweder die Identifizierung von für das Ermittlungsverfahren relevante Datensätzen, die reale Objekte oder Personen repräsentieren oder das Auffinden von Inhalten, die entweder Tatobjekt (z. B. Texte im Bereich von Propagandadelikten) oder andere inhaltliche Spuren, die für eine Beweisführung (z. B. E-Mail mit tatrelevanten Inhalten wie z. B. Tatabsprachen) herangezogen werden können.

Dabei können verschiedene Analysemethoden zum Einsatz kommen. Nachfolgend werden exemplarisch vier wesentliche Analysemethoden vorgestellt.

4.5.1 Soziale Netzwerkanalyse

Eine für Ermittlungsbehörden attraktive Datenanalysevariante im Big Data-Bereich ist die soziale Netzwerkanalyse (englisch: Social Network Analysis; kurz SNA).[300] Aber was wird unter einem sozialen Netzwerk verstanden? Unter einem sozialen Netzwerk versteht Lothar Krempel[301] die Beziehungen zwischen verschiedenen Akteuren, die unter anderem Personen, Firmen, Organisationen oder auch Nationen sein können. Beziehungen zwischen diesen Akteuren können z. B. Kontakte, Mitgliedschaften oder Austauschprozesse sein.

Für Johannes Weyer können soziale Netzwerke als

> „[. . .] eine eigenständige Form der Koordination von Interaktionen verstanden werden, deren Kern die *vertrauensvolle Kooperation* [Hervorhebung im Original; Anm. des Autors] autonomer, aber interdependenter (wechselseitig voneinander abhängiger) Akteure ist, die für einen begrenzten Zeitraum zusammenarbeiten und dabei auf die Interessen des jeweiligen Partners Rücksicht nehmen, weil sie auf diese Weise ihre partikularen Ziele besser realisieren können als durch nichtkoordiniertes Handeln.“[302]

[300] Vgl. WOJCIECHOWSKI 2012, S. 52.
[301] Vgl. KREMPEL 2005, S. 29.
[302] WEYER 2011b, S. 49.

Maßgeblich ist, dass die Akteure durch diese Netzwerke verbunden sind. Beziehungsgeflechte entstehen bei Menschen aufgrund „[. . .] ihrer elementaren Ausgerichtetheit, ihrer Angewiesenheit aufeinander und ihrer Abhängigkeit voneinander [. . .]“[303] Die Feststellungen von Elias und Weyer lassen sich in zwei verschiedene Ausprägungen ins kriminalistische Umfeld übertragen: Zum einen sind Täter und Opfer – wie alle anderen Personen auch – in alltägliche soziale Geflechte eingebunden. Zudem können Täter aufgrund einer gemeinsamen Zielrichtung innerhalb von Tätergruppierungen, dem arbeitsteiligen Vorgehen und den Abhängigkeiten (z. B. beim Erwerb von Tatmitteln oder beim Absetzen der Tatbeute) zusätzliche Interdependenzgeflechte bilden, die neben den üblichen legalen Netzwerken bestehen. Damit umfasst der Begriff des sozialen Netzwerkes auch Beziehungen zwischen Tätern, ihren Unterstützern, ihren Opfern aber auch ihres sozialen Umfeldes, bestehend aus ihrer Familie, Freunden und auch Bekannten, die keinen Bezug zur Straftat haben müssen. Die Repräsentation dieser Beziehungen ist dabei nicht nur auf den direkten Austausch zwischen den Einzelnen beschränkt, sondern kann auch durch Austauschprozesse dargestellt werden, die durch die Akteure angestoßen werden.[304] Hierzu zählt u. a. eine direkte Kommunikation wie auch z. B. der Transfer von Waren und/oder Geld.

Ein soziales Netzwerk wird durch Graphen repräsentiert,[305] welche ein mathematisches Modell für netzartige Strukturen in der Natur und Technik darstellen. Diese bestehen aus zwei verschiedenartigen Objekten, nämlich Knoten (Akteure) und Kanten (Beziehungen).[306] Innerhalb dieser Graphen können verschiedene Zustände – gerichtet und ungerichtet – im sozialen Netzwerk dargestellt werden. **Gerichtete Graphen** zeichnen sich dadurch aus, dass die Wirkungsrichtung der Kanten zwischen zwei Knoten beschrieben ist. So kann in einem sozialen Netzwerk dargestellt werden, dass z. B. Geld vom Konto des Akteurs A zum Konto des Akteurs B geflossen ist oder dass Akteur B den Akteur A anrief. Wirkungsrichtungen können auch wechselseitig erfolgen. Wird die Richtung der Relation nicht unterschieden, so spricht man von einem **ungerichteten Graphen**.[307] Ein weiteres Unterscheidungsmerkmal ist die Stärke von Beziehungen.[308] Diese Gewichtung spiegelt sich z. B. in der Häufigkeit von telefonischen Kontakten oder der Summe eines oder mehrerer Geldtransfers wider. An dieser Stelle wird aber auch deutlich, dass die Definition der Gewichtung von einer zugrundeliegenden Hypothese abhängig ist, in der z. B. die Häufigkeit eines Kontaktes oder die Überweisungshöhe eine Aussage über die Stärke einer Beziehung zwischen zwei Akteuren zulässt.

[303] ELIAS 2009, S. 12.
[304] Vgl. KREMPEL 2005, S. 29.
[305] Vgl. KREMPEL 2005, S. 73 f.
[306] Vgl. TITTMANN 2011, S. 11.
[307] Vgl. KREMPEL 2005, S. 80.
[308] Vgl. STEGBAUER/RAUSCH 2013, S. 26.

Damit wird erkennbar, dass die Darstellung von Zusammenhängen innerhalb eines sozialen Netzwerkes erst einmal nur ein analytisches Konstrukt ist, welches keine Aussage zulässt, ob tatsächlich die so konturierten Zusammenhänge in der Realität bestanden.[309] Diese konstruierten Zusammenhänge gilt es später mit anderen Mitteln – z. B. Vernehmungen – zu verifizieren.

Ein soziales Netzwerk kann anhand verschiedener Kriterien analysiert werden. Kriminalistisches Ziel einer solchen Analyse können z. B. die Verdachtsgewinnung bei Betrugssachverhalten[310], die Aufhellung von Personen- oder Organisationsstrukturen und die Kommunikationsstrukturen einer technischen Infrastruktur im Bereich der Cybercrime[311] sein. Dabei ist die SNA eine statische Momentaufnahme einer Konstellation zu einem bestimmten Zeitpunkt, der in der Vergangenheit liegt.[312]

Die Analyse kann auf unterschiedlichen Wegen erfolgen. Eine erste Möglichkeit ist die Darstellung des sozialen Netzwerkes als grafische Visualisierung als Netz (siehe Abbildung 7). Hierdurch wird der Bearbeiter in die Lage versetzt, im Rahmen von Visual Analytics (siehe Kapitel 2.4.3) eigene Schlüsse über das Netzwerk zu ziehen.

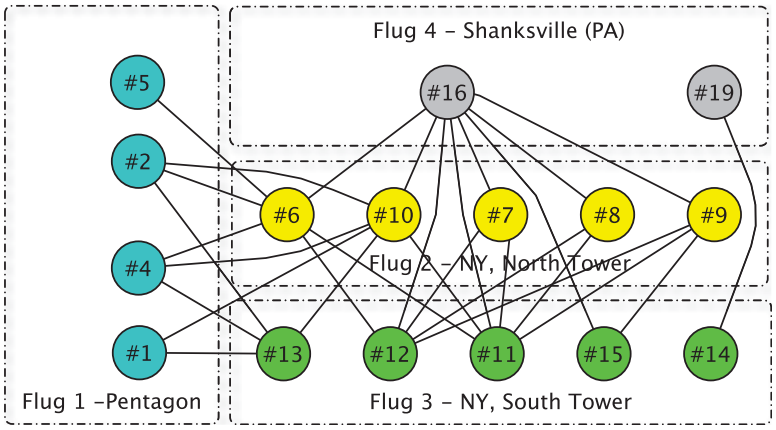

Abbildung 7: Beispiel eines sozialen Netzwerkes[313]

Eine weitere Analyseform ist die Bestimmung der Positionen der Akteure in Netzwerken durch Berechnung einer Rangfolge aufgrund verschiedener Kriterien.[314] Zur Bestimmung der Rangfolge der einzelnen Akteure in einem sozialen Netzwerk werden den Knoten (Akteure) durch Berechnungen verschiedene Kennzahlen zugewiesen. Dabei handelt es sich um die Zentralitätsmaße degree, closeness und

[309] Vgl. WEYER/FINK/LIBSCHIK 2011, S. 117.
[310] Vgl. WEYER 2011a, S. 13.
[311] Vgl. BUNDESKRIMINALAMT 2015, S. 69.
[312] Vgl. WEYER/FINK/LIBSCHIK 2011, S. 118.
[313] Abbildung 7: Eigene Darstellung. Beziehungen zwischen den Attentätern vom 11.09.2001 (eingeschränkt auf die Akteure, die mindestens zu einem weiteren Attentäter eines anderen Flugs Kontakt hatte). Ein ausführlicheres Beispiel – inkl. Rohdaten und Zentralitätsmaßen – befindet sich in Anlage D.
[314] Vgl. WEYER/FINK/LIBSCHIK 2011, S. 117 f.

betweeness. Der Grad **degree** beschreibt die Anzahl der direkten Beziehungen, mit denen ein Knoten verbunden ist. Ein Akteur (Knoten), der andere Akteure im Netz durch viele direkte Beziehungen (Kanten) erreichen kann, gilt als zentral (degree centrality).[315] Dorothea Jansen und Rainer Diaz-Bone[316] beschreiben, dass ein hoher Degree für ein hohes soziales Kapital spricht. Sie verstehen unter sozialem Kapital „einen vorteilhaften Effekt der Netzwerkstruktur, der darin besteht, individuellen oder korporativen Akteuren breitere Handlungsmöglichkeiten oder Zugang zu Ressourcen zu eröffnen.“[317] Das zweite Zentralitätsmaß – die **closeness**-Zentralität – berücksichtigt neben den direkten auch die indirekten Verbindungen eines Akteurs zu anderen Akteuren. Dabei gilt ein Knoten als zentral, wenn er alle anderen Knoten auf sehr kurze Weise erreichen kann. Dazu wird die Gesamtdistanz der kürzesten Pfade zwischen diesen und denen anderer Akteure berechnet.[318] Das dritte Zentralitätsmaß ist die **betweeness**-Zentralität, die angibt, wie oft ein Knoten als Makler auf dem kürzesten Weg zwischen den anderen Akteuren liegt.[319]

Betrachtet man unter Berücksichtigung der eben beschriebenen Netzwerkpositionen eine Organisation als Netzwerk, so kann geschlussfolgert werden, dass ein Akteur mit hoher degree-Zentralität einen direkten Einfluss auf eine große Anzahl von Mitarbeitern hat. Ein Akteur mit hoher closeness-Zentralität vermag dies ohne direkten Kontakt über wenige Mittelsmänner, wie z. B. Führungskräfte in einer niedrigeren Hierarchiestufe erreichen. Der Akteur mit einem hohen betweeness-Zentralitätsmaß hat Zugriff auf ein hohes Informationsaufkommen und könnte so Einfluss auf Informationsflüsse und Entscheidungsprozesse ausüben. Daher sind diese Werte gute Anhaltspunkte für eine erste Bewertung von Positionen einzelner Personen in einem Netzwerk.

Jedoch muss beachtet werden, dass Verzerrungen auftreten können, die eine Zentralität einzelner Personen suggeriert, die aber nur aufgrund der fokussierten Betrachtung einzelner Personen entsteht, da überwiegend ihn betreffende Daten in die Auswertung einfließen.[320] Dieser Umstand ist bei der Bewertung der Ergebnisse zu beachten und muss in die Erstellung entsprechender Hypothesen einfließen.

Nun stellt sich noch die Frage, wie die zur Erstellung eines solchen Netzwerkes erforderlichen Daten erhoben werden können. In der Sozialwissenschaft erfolgt die Datenerhebung z. B. durch Interviews oder Beobachtung.[321] Im kriminalistischen Umfeld können ebenfalls Daten zu einem sozialen Netzwerk mit Hilfe von kriminalistischen Interviews, also im Rahmen von Vernehmungen, oder kriminalistischen

[315] Vgl. KREMPEL 2005, S. 130.
[316] Vgl. JANSEN/DIAZ-BONE 2011, S. 86.
[317] JANSEN/DIAZ-BONE 2011, S. 75.
[318] Vgl. KREMPEL 2005, S. 76.
[319] Vgl. JANSEN/DIAZ-BONE 2011, S. 101.
[320] Vgl. NITSCHKE/ROLLINGER 2015, S. 235.
[321] Vgl. HÄUSSLING 2010, S. 79.

Beobachtungen, im Rahmen von Observationen, erhoben werden. Ebenso besteht die Möglichkeit, Netzwerkinformationen im Rahmen einer Big Data-Datenerhebung für eine soziale Netzwerkanalyse zugänglich zu machen.

Die Erstellung sozialer Netzwerke ist z. B. über die Analyse von E-Mail- oder Telefonverbindungsdaten möglich. In einer Studie der Universitäten Leuven und Rotterdam wurden die E-Mailkontakte von 2.300 Personen, die in einem Zeitraum von drei Jahren angefallen sind, untersucht. Dabei wurde festgestellt, dass für eine vollständige Aufdeckung aller Netzwerkbeziehungen lediglich die Daten von acht Prozent der Gruppe hätten ausgewertet werden müssen.[322] Zum gleichen Ergebnis kam ein Team des Max-Planck-Instituts, das im Auftrag des Bundesministeriums der Justiz forschte. Dabei stellte Hans-Jörg Albrecht et al. im Jahr 2008 fest, dass die Überwachung von acht Prozent der Beteiligten einer kleinen Gruppe ausreicht, um ein Netzwerk vollständig aufzudecken.[323] Zudem kam eine Studie des Massachusetts Institute of Technology und der Harvard Universität[324] zur Erkenntnis, dass die Auswertung der Mobilfunk- und Bluetooth-Verkehrsdaten geeignet waren, um 95% aller Freundschaftsbeziehungen festzustellen. Aus diesen Forschungsergebnissen kann abgeleitet werden, dass die Erhebung von Verkehrsdaten (Telefon und E-Mail) zur Schaffung einer Datenbasis zur Netzwerkanalyse geeignet ist.

Für die Analyse von sozialen Netzwerken sind verschiedene Softwareprodukte auf dem Markt vorhanden. So bietet die Firma IBM die kommerzielle Software i2 Analyst's Notebook an, welche explizit zur Auswertung sozialer Netzwerke u. a. für polizeiliche Bedarfsträger entwickelt wurde.[325] Diese Software ermöglicht die Erstellung von Visualisierungen von Netzwerken wie auch die Berechnung der dargestellten Zentralitätswerte. Ebenso gibt es verschiedene Open Source und andere (freie) Software wie z. B. die Software yED[326], NetDraw[327] oder Gephy[328]. Diese Softwareprodukte erlauben die im ETL-Prozess aufbereiteten Daten zu importieren und nach den oben genannten Fragestellungen (Zentralitätsmaß) zu analysieren.

Als Hilfsmittel kommen aber nicht nur vollständige Softwareprodukte zum Einsatz. Auch mit Hilfe von Programmiersprachen wie z. B. R[329], welche auf die statistische Auswertung von großen Datenmengen spezialisiert ist, können unter Einbindung entsprechender Erweiterungen soziale Netzwerke ausgewertet und visualisiert werden.

Wie sich gezeigt hat, ermöglicht die soziale Netzwerkanalyse auch in der Krimi-

[322] Vgl. SCHULZKI-HADDOUTI 2011, S. 38.
[323] Vgl. ALBRECHT/GRAFE/KILCHLING 2008, S. 106.
[324] Vgl. EAGLE/PENTLAND/LAZER 2009.
[325] Vgl. IBM 2012.
[326] https://www.yworks.com/de/products_yed_about.html (besucht am 03.12.2017).
[327] https://sites.google.com/site/netdrawsoftware/home (besucht am 03.12.2017).
[328] https://gephi.org (besucht am 03.12.2017).
[329] R ist eine freie Softwareumgebung für statistische Berechnungen. Für R werden Erweiterungspakete für die SNA angeboten. https://www.r-project.org/ (besucht am 03.12.2017).

nalistik, Beziehungen zwischen Menschen zu analysieren. Die benötigten Daten werden zuvor im ETL-Prozess aufbereitet, sodass sie in die Analyseanwendung geladen werden können. In dieser Anwendung können sie entweder visualisiert werden oder die Daten dienen als Grundlage für die Berechnung von Zentralitätskennzahlen der Akteure des Netzwerks, die ihnen eine konkrete Position im Netzwerk zuschreiben. Das Ergebnis stellt sich als Hypothese dar, dass z. B. eine bestimmte Person eine besondere Stellung in einer Gruppe einnimmt. Diese Hypothese wird datengetrieben generiert, sodass diese Methode gemäß der unter Kapitel 2.4.1 beschriebenen Definition als Data Mining betrachtet werden kann.

Die soziale Netzwerkanalyse kann mithin in Ermittlungsverfahren eingesetzt werden, wenn in der Zieldefinition (Kapitel 4.1) Hypothesen und Ermittlungsfragen zur Funktion von Personen in Netzwerken und Organisationen aufgestellt werden.

4.5.2 Mengenanalyse

Eine weitere Herangehensweise, aus großen Datenmengen die Datensätze herauszufiltern, die für die weitere Ermittlung von Bedeutung sind, ist die Mengenanalyse. In diesem Zusammenhang wird oft von Schnittmengen gesprochen, die aus den Datenbeständen herausgearbeitet werden sollen.[330] Diese Bezeichnung beschreibt – wie sich noch zeigen wird – diese Analyseform nicht vollständig, da nicht nur Schnittmengen, sondern auch Ausschluss-, Differenz- und Vereinigungsmengen als Grundlage für die Selektion der relevanten Daten dienen. Daher wird nachfolgend der Oberbegriff Mengenanalyse genutzt.

Bei einer Mengenanalyse werden verschiedene, für die Analyse erhobene Datenbestände, die die gleichen Datenobjekte (z. B. Personen) enthalten, gegenübergestellt und anhand der durch für die Klärung der Hypothesen operationalisierten Parameter reduziert. Dabei werden die einzelnen Datenbestände als Filterkriterium des anderen Datenbestandes genutzt.

Diese Datenanalysevariante soll beispielhaft anhand der Auswertung von Rasterfahndungs- und von Funkzellendaten sowie des Datenabgleichs erläutert werden.

Beispiel: Rasterfahndung

Die Rasterfahndung wird im § 98 a Abs. 1 Satz 1 StPO geregelt. Danach dürfen

> „[. . .]personenbezogene Daten von Personen, die bestimmte, auf den Täter vermutlich zutreffende Prüfungsmerkmale erfüllen, mit anderen Daten maschinell abgeglichen werden, um Nichtverdächtige auszuschließen oder Personen festzustellen, die weitere für die Ermittlungen bedeutsame Prüfungsmerkmale erfüllen[. . .]“

[330] Vgl. LUDWIG 2008, S. 254.

Die Formulierung macht deutlich, dass aus einer Vielzahl von Daten die Datenmenge extrahiert werden soll, die einem bestimmten Merkmalsschema des Täters entspricht. Ausgangspunkt bei einer Rasterfahndung sind eine oder mehrere Hypothesen, die sich auf die Person des Täters und auf Merkmale seines Handelns beziehen. Diese Merkmale müssen direkt oder indirekt als Daten in Datenbanken gespeichert sein. Die für die Rasterfahndung erforderlichen Täterhypothesen werden im Schritt 1 des kriminalistischen Big Data-Auswertungszyklus im Rahmen der Fallanalyse erstellt (siehe Kapitel 4.1). Von diesen Hypothesen werden dann die erforderlichen Attribute und der zu erhebende Datenbestand abgeleitet. Ein ausführliches Beispiel erläuterte 1986 der ehemalige Präsident des Bundeskriminalamtes Horst Herold in einem Interview mit dem Spiegel. In diesem Interview beschreibt er, wie individuelle Erkenntnisse zu Tatverdächtigen die Datenmenge von polizeiexternen Datenbeständen sukzessive reduzieren können, bis zum Schluss die Daten übrigbleiben, die den oder die Täter repräsentieren. Dabei handelte es sich um die Rastermerkmale (hier ein Auszug), die vom BKA für die Fahndung nach RAF-Terroristen herausgearbeitet wurden:[331]

1. Wohnungsmieter
2. Stromrechnung bar bezahlt
3. Keine Personen mit legalen Namen
    a) nicht als Einwohner gemeldet,
    b) kein Kfz-Halter,
    c) kein Rentner,
    d) kein Bafög-Bezieher,
    e) nicht im Grundbuch als Eigentümer verzeichnet,
    f) nicht brandversichert,
    g) nicht gesetzlich krankenversichert

Bei einen solchen Abgleich von Daten werden verschiedene Methoden der Mengenlehre angewendet. Beim Abgleich der Daten der Gruppe 1 mit der Gruppe 2 wird eine Schnittmenge der Personen von Wohnungsmietern und den Personen, die ihre Stromrechnung bar zahlen, gebildet. Im zweiten Schritt wird eine Vereinigungsmenge von Personen gebildet, die gemäß einer Hypothese als Person mit legalen Namen gelten. So werden die Datenbestände der Einwohnermeldeämter, Kfz-Zulassungsstellen, Bafög-Ämter, Grundbuchämter und Versicherungen zu einem Datenpool verschmolzen. Im dritten Schritt wird die Differenzmenge gebildet, in dem von der Schnittmenge der Gruppe 1 und Gruppe 2 die Daten der Vereinigungsmenge der Gruppe 3 ausgeschlossen werden.

Im Falle der von Herold geschilderten Rasterfahndung im Zusammenhang mit der RAF konnten durch diese Methode zwei Personen ermittelt werden, nämlich

[331] Vgl. HENTSCHEL/PÖTZL 1986, S. 49.

ein Rauschgifthändler und der RAF-Terrorist Rolf Heißler.[332]

Beispiel: Funkzellenauswertung

Die Mengenanalyse wird ebenfalls im Rahmen erhobener Funkzellendaten genutzt.

> „Eine [...] **Funkzellenauswertung** [Hervorhebung im Original; Anm. des Autors] ist eine kriminalistische Maßnahme zur Eingrenzung stattgefundener Telekommunikation in einem näher bezeichneten räumlichen und zeitlichen Sektor."[333]

Wie bei der Rasterfahndung soll bei der Funkzellenauswertung der Datenbestand auf die relevante Telekommunikation eingegrenzt werden. Dabei wird auf den Datenbestand der Mobilfunkbetreiber zurückgegriffen, der grundsätzlich die Rufnummer in der Funkzelle, die Partnernummer der Telekommunikation (in oder außerhalb der Funkzelle), die Angabe zum Telekommunikationsdienst (Gespräch, Internet, SMS), Informationen zur IMSI[334] und IMEI[335] und den Zeitpunkt der Kommunikation enthält.[336]

Es gibt zwei grundsätzlich verschiedene Ausgangssituationen, in denen eine Funkzellenauswertung eingesetzt werden kann. Zum einem kann eine Einzeltat vorliegen, bei der aufgrund des Modus Operandi die Hypothese zulässig ist, dass der Täter an einem oder mehreren bestimmbaren Orten (z. B. Tatvorbereitungsort, Tatort, Fluchtweg) zu einem bestimmbaren Zeitpunkt einen Telekommunikationsdienst genutzt hat und durch eine Analyse der Funkzellendaten diese Telekommunikationshandlung selektiert werden kann. Zum anderen kann eine Serienstraftat vorliegen, bei der die Hypothese aufgestellt werden kann, dass der Täter an den verschiedenen Tatorten sein Mobiltelefon genutzt hat, dessen Identifikationsmerkmale in den Datenbeständen durch eine Analyse gefunden werden kann.[337] Diese beiden unterschiedlichen Ausgangssituationen bedürfen der gleichen Vorgehensweise.

In beiden Anwendungsfällen werden die Funkzellendaten der verschiedenen Handlungsorte des Täters oder des gleichen Ortes zu verschiedenen Zeitpunkten (Tatphasen) geschnitten.[338] Dabei muss die gültige Hypothese bestehen, dass in den verschiedenen Datenbeständen zum gleichen Telekommunikationsmittel entsprechende Identifikationsmerkmal, wie z. B. Rufnummer, IMSI oder IMEI, enthal-

[332] Vgl. BURGHARD 1993, S. 72 f.
[333] HEINRICHS/WILHELM 2010.
[334] Bei der International Mobile Subscriber Identity (IMSI) handelt es sich um die auf der SIM-Karte gespeicherte Teilnehmeridentifikationsnummer; vgl. ALBRECHT/GRAFE/KILCHLING 2008, S. 45.
[335] Bei der International Mobile Equipment Identification (IMEI) handelt es sich um die elektronische Gerätekennung des mobilen Endgerätes; vgl. ALBRECHT/GRAFE/KILCHLING 2008, S. 45.
[336] Vgl. HEINRICHS/WILHELM 2010; Vgl. HABERBERGER/TALARCZYK 2007, S. 369.
[337] Vgl. HEINRICHS/WILHELM 2010.
[338] Vgl. ROERICH 2017, S. 176.

ten sind. Zudem muss die konkrete Annahme bestehen, dass der Täter aufgrund seiner Persönlichkeit oder der Tatbegehungsweise an verschiedenen bestimmbaren Orten mittels des gleichen Endgerätes oder verschiedenen Endgeräten mit der gleichen SIM-Karte kommuniziert hat.

Eine solche Möglichkeit könnte vorliegen, wenn z. B. Mitglieder einer Tätergruppierung zu Absprachezwecken bei mehreren Straftaten am jeweiligen Tatort mit anderen Mitgliedern der Gruppen außerhalb und innerhalb der gleichen Funkzelle telefonieren müssen. Diese Täterkommunikation kann – bezogen auf die Funkzellenauswertung – verschiedene Erscheinungsformen annehmen, die es bei der Funkzellenanalyse abzuklären gilt:

1. Das gleiche Endgerät befindet sich an verschiedenen Zeitpunkten in der Funkzelle verschiedener kriminalistisch relevanter Orte (z. B. Tatorte oder Annäherungs- und Fluchtwege) bzw. das gleiche Endgerät befindet sich zu verschiedenen Zeitpunkten in der gleichen Funkzelle, wenn der Ort gleich ist

2. Unterschiedliche Endgeräte stehen in verschiedenen Funkzellen mit dem gleichen Telekommunikationspartner (z. B. Koordinator) in Verbindung

3. Unterschiedliche Endgeräte stehen zur gleichen Zeit in den Funkzellen verschiedener kriminalistisch relevanter Orte in Verbindung

Zur Identifizierung der verdächtigen Mobilfunkgeräte werden die Telekommunikationsdaten von mindestens zwei verschiedenen Orten oder an einem Ort aber zu mindestens verschiedenen Zeiten abgeglichen.

Durch einen Datenmengenvergleich können Rufnummern, Geräte- oder Kartennummern als Schnittmenge ermittelt werden, die einer oder mehrere der oben aufgeführten Hypothesen entsprechen. Das Ergebnis kann jedoch verschiedene fachliche Störfaktoren enthalten. So können z. B. Servicerufnummern (z. B. Essenslieferanten, Störungsstellen der Mobilfunkbetreiber) oder Rufnummern von Personen enthalten sein, die sich in einem bestimmten Bereich berechtigt aufhalten und auch mobil kommunizieren (z. B. Anwohner). Je näher die räumliche Entfernung der Funkzellen zu einander ist, umso höher sind diese Anwohnertreffer.[339] Daher muss die Treffermenge so lange bereinigt werden, bis eine Menge an Rufnummern übrigbleibt, die durch klassische kriminalistische Maßnahmen (z. B. Vernehmungen oder Alibiermittlungen) ausermittelt werden können.

Der Kreis der berechtigten Personen kann hierfür im Vorfeld festgelegt werden. Dazu können automatisierte Verfahren genutzt werden, die den Datenbestand zur Generierung der Ausschlussmenge erstellt (z. B. durch automatisierte Zuordnung von Rufnummer $\mapsto$ Anschlussinhaber $\mapsto$ Wohnanschrift $\mapsto$ Versorgungsgebiet der Funkzelle). Zudem können bereits im ETL-Prozess die Funkzellendaten

[339] Vgl. HEINRICHS/WILHELM 2010.

über Datensammlungen von Servicerufnummern angereichert werden, um diese dann vom Ergebnis auszuschließen. Ein weiterer Störfaktor bei der Auswertung von Serientaten über einen längeren Zeitraum ist, dass Täter ihre Mobiltelefone oder ihre Rufnummern wechseln können.[340] Eine Ermittlung von Kartenwechslern ist aufgrund der mitgelieferten IMEI- und IMSI-Daten möglich, da diese das Mobiltelefon bzw. die SIM-Karte identifizieren.

Beispiel: Datenabgleich mit polizeilichen Daten

Das dritte Anwendungsbeispiel für die Nutzung des Mengenabgleiches ist der Datenabgleich von polizeilichen Datenbanken. Dieser Datenabgleich ist in § 98 c StPO geregelt und erlaubt den maschinellen Abgleich von personenbezogenen Daten mit anderen zur Strafverfolgung, Strafvollstreckung oder zur Gefahrenabwehr gespeicherten Daten. Durch diese Regelung soll klargestellt werden, dass der Datenabgleich nur mit den Datenbeständen durchgeführt werden darf, die den Ermittlungsbehörden bereits aufgrund ihrer Tätigkeit (auch aus Maßnahmen der Gefahrenabwehr oder anderer Strafverfahren) vorliegen.[341] Beim Datenabgleich im Rahmen der Big Data-Analyse handelt es sich systematisch ebenfalls um einen Mengenabgleich, bei dem die Schnittmenge von der anfragenden Datenmenge und der Datenmenge des polizeilichen Datenbestandes gebildet wird.

Im Rahmen eines solchen Abgleichs könnten z. B. Rufnummernlisten eines sichergestellten Mobiltelefons mit den Rufnummern geschnitten werden, die im Rahmen des Strafverfahrens ebenfalls bereits erhoben und z. B. in einem Fallbearbeitungssystem oder der TKÜ-Anwendung gespeichert wurden.

Zweck des Abgleiches ist nach Auffassung von Clages[342] die Gewinnung von Informationen, die einen strafprozessualen Verdacht begründen oder verdichten können bzw. Anhaltspunkte für weitere Ermittlungsmaßnahmen erbringen. Grundsätzlich beantwortet ein Abgleich nur die Frage, ob eine zu überprüfende Entität bereits im Datenbestand der Ermittlungsbehörde gespeichert ist. Ein Nichttreffer im Verfahrensdatenbestand erzeugt für das Ermittlungsverfahren nur die Erkenntnis, dass diese Information bisher nicht gespeichert war. Bei einem Treffer ergeben sich aus den Metadaten der Anfrage- und Trefferdaten und der konkreten Trefferkombination verschiedene Szenarien der Informationsgewinnung (siehe Abbildung 8). Bei einem Treffer im Datenbestand des konkreten Ermittlungsverfahrens kann z. B. die Relevanzeinstufung dieser Daten für den anfragenden Datenbestand übernommen werden. So kann ein Mobiltelefon für die Bearbeitung priorisiert werden, wenn in seinem Rufnummernspeicher eine Telefonnummer gespeichert ist, die in einer TKÜ aufgrund von verfahrensrelevanten Gesprächen aufgefallen ist.

[340] Vgl. LUDWIG 2008, S. 256.
[341] Vgl. KÖRFFER 2014, S. 148.
[342] Vgl. CLAGES 2017c, S. 287.

Ein weiterer Mehrwert entsteht, wenn durch einen Treffer neue Informationen aus einem anderen Verfahren oder Vorgang gewonnen werden können. Diese können neue Anhaltspunkte, z. B. bisher im Verfahren nicht bekannte Personenkontakte oder Rufnummern, erbringen, die wiederum neue Ermittlungsansätze für das konkrete Ermittlungsverfahren bieten. Umgekehrt können auch bereits vorliegende Informationen um neue Erkenntnisse aus den neuen Daten ergänzt werden.

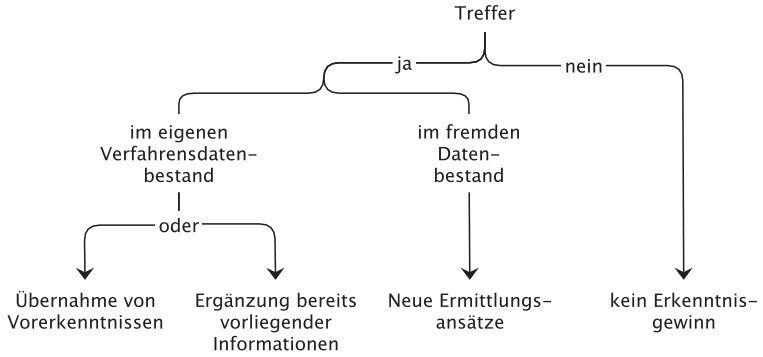

Abbildung 8: Erkenntnisgewinn aus Datenabgleichen[343]

Im Umfeld des Datenabgleichs gibt es jedoch eine Vielzahl von Einschränkungen und Hindernisse, die insbesondere beim Abgleich von Personennamen auftreten können. So können z. B. Transkriptionen[344] und abgeleitete Namensweisen wie Verniedlichungen[345] den Abgleich erschweren. Diese Umstände können sogar dazu führen, dass ein Datensatz nicht gefunden wird, obwohl die repräsentierte Person in den Datenbeständen enthalten war. Ebenso ist die Datenqualität von entscheidender Rolle.

Generelle Vorgehensweise bei einer Mengenanalyse

Bei der Selektion von Daten aus einer großen Datenmenge können weitere Datenmengen als Filterkriterien herangezogen werden. Durch das Erzeugen von Schnittmengen werden die Daten identifiziert, die in beiden Datenmengen enthalten sind. Weiterhin können Datenmengen genutzt werden, um aus der Datenmenge mittels Differenzmengenbildung nicht relevante Daten auszuschließen. Diese Bildung von Schnitt- und Differenzmengen kann Schritt für Schritt mit jedem vorliegenden Datenbestand durchgeführt werden. Alternativ besteht die Möglichkeit, die Datenbestände im Vorfeld zu einer Vereinigungsmenge zusammenzufassen.

Die Aussagekraft dieser Analysemethode ist dabei stark abhängig von den Datenbeständen und dem Informationsgehalt dieser Daten, der sich durch die Kombination fallabhängig verändert. Während z. B. in einem Fall ein Nichttreffer ein stark

[343] Abbilung 8: eigene Darstellung.
[344] Vgl. LISBACH 2011, S. 41 ff.
[345] Vgl. LISBACH 2011, S. 61 ff.

entlastendes Indiz ist (RAF-Beispiel), kann es in einem anderen Fall kaum eine entlastende Wirkung entfalten (Datenabgleichsbeispiel). Zudem ist die Interpretation von Treffern und Nichttreffern von weiteren Komponenten wie Zeit und Ort der Tat (Funkzellenbeispiel) abhängig. Es ist also wichtig, dass die Ausgangsdaten, die Ergebnisse und die Rahmenbedingungen zur Interpretationsunterstützung der Ergebnisse dokumentiert sind.

4.5.3 Geodatenanalyse

Der dritte Typ der kriminalistischen Wertschöpfung aus Big Data ist die Analyse von geografischen Daten, auch als Geoinformationen bezeichnet. Ausgangspunkt der geografischen Auswertung ist wiederum die Erstellung einer oder mehrerer Hypothesen im ersten Prozessschritt (siehe Kapitel 4.1) des kriminalistischen Big Data-Auswertungszyklus'. Eine zwingende Voraussetzung ist hier, dass die Hypothesen einen Raumbezug enthalten.

Die Auswertung von geografischen Daten aus Straftaten kann grundsätzlich in zwei Zielrichtungen erfolgen. So können zum einen die Daten zur Vorbeugung von Straftaten im Rahmen des Predictive Policing analysiert werden. Diese Methode ist in die Zukunft gerichtet, auch wenn hierfür Daten bereits verübter Straftaten in Verbindung mit weiteren georeferenzierten Daten[346] ausgewertet werden. Zum anderen können geografische Daten zur Aufarbeitung eines kriminalistisch relevanten Ereignisses ausgewertet werden. In der nachfolgenden Betrachtung soll beispielhaft dargestellt werden, welche Informationen mit Hilfe der geografischen Analyse von Big Data für die Fallaufklärung gewonnen werden können.

Geografische Daten können in Ermittlungsverfahren aus einer Vielzahl von Quellen gewonnen werden. Hierfür kommen in erster Linie Geodaten in Betracht, die im Zusammenhang mit der Übermittlung der (rückwirkenden) Verbindungsdaten von den Providern übermittelt werden.[347] Zudem werden Geopositionsdaten durch Nutzung von stillen SMS[348] erzeugt, während einer TKÜ-Maßnahme[349] oder bei Observationsmaßnahmen[350] durch die Polizei aufgezeichnet und können aus Datenbanken aus sichergestellten Mobiltelefonen extrahiert werden.[351] Zudem können Geodaten auch nachträglich durch die Polizei zu Dokumentationszwecken festgehalten werden, z. B. wenn im Rahmen eines Mantrailing Fluchtwege des

[346] Vgl. ROLFES 2017, S. 57.
[347] Vgl. HABERBERGER/TALARCZYK 2007, S. 369.
[348] Unter ***stille SMS*** wird der Versand von Ortungsimpulsen zur Bestimmung des Standortes einer Mobilfunk-Basisstation, welche die gegenwärtige vorsorgende Funkzelle für ein überwachtes mobiles Endgerät aufspannt, verstanden. vgl. auch BT-Drs. 18/2932, 10 f.
[349] Vgl. BUNDESNETZAGENTUR FÜR ELEKTRIZITÄT, GAS, TELEKOMMUNIKATION, POST UND EISENBAHNEN 2017, S. 50.
[350] Vgl. MEYER-GOSSNER/SCHMITT 2015, S. 427.
[351] Vgl. JÜNGLING 2011.

Täters durch einen Polizeispürhund abgelaufen werden.[352] Neben diesen direkt erfassten geografischen Daten, die als Geopositionsdaten mit Längen- und Breitengrad übermittelt werden[353], liegen oftmals auch Geodaten in indirekter Form vor. Eine indirekte Geoinformation besteht dann, wenn einem Datenobjekt, welches selbst nicht über Geoinformationen verfügt, durch eine Relation über mehrere Ebenen zu einer geografischen Position ein Ort zugeordnet werden kann. So kann z. B. die Auswertung von Geldautomatenabhebungen den Aufenthaltsort eines Geldkartenbesitzers zu einem bestimmten Zeitpunkt bestimmen. Hier werden Nutzer mit den Standorten der Geldautomaten in Verbindung gesetzt, sodass nun der Standort einer bestimmten Person festgestellt ist (siehe Abbildung 9). Über diesen Weg können eine Vielzahl von Daten um eine geographische Position ergänzt werden, die dann im Rahmen einer geografischen Auswertung genutzt werden können. Aus diesem Beispiel wird noch eine weitere Komponente erkennbar,

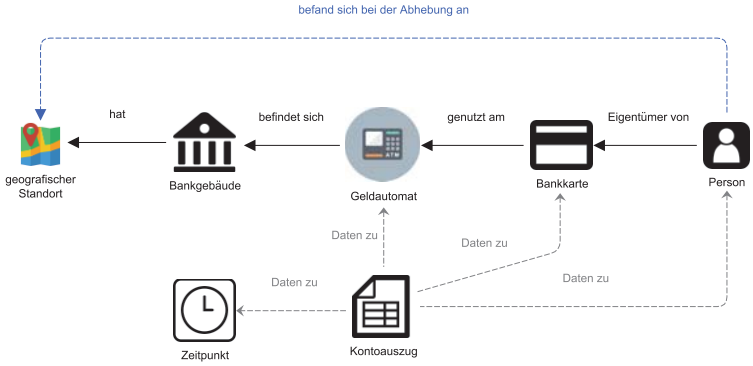

Abbildung 9: Zuordnung eines Aufenthalts einer Person mit Hilfe indirekter Geopositionsdaten[354]

die für die kriminalistische Auswertung von geografischen Daten wichtig ist. Neben den konkreten Ortsdaten in direkter oder indirekter Form wird der Zeitpunkt benötigt, um einen Wert zur Sachverhaltsaufklärung zu generieren. Überwiegend ist es nicht nur entscheidend, an welchem bestimmten Ort eine Person war, sondern auch wann sie sich dort aufgehalten hat. Dabei ist die Nutzung von Ort- und Zeitangaben keine neue Analysemöglichkeit, die erst mit der Einführung von Geoinformationssystemen Einzug in die Ermittlungsarbeit erhalten hat. Die Erstellung von Weg-Zeit-Diagrammen ist eine anerkannte und seit langem praktizierte Methode, auch ohne die Unterstützung von Computern.[355] Dass die Zeitangabe relevant ist, lässt sich an folgendem Beispiel leicht ableiten: Ein Treffen von zwei Personen liegt nur dann vor, wenn sie nicht nur räumlich am selben Ort waren, sondern auch zeitlich. Daher ist der Qualität der Zeitangaben im Raumkontext ein besonderer Wert

[352] Vgl. BUDDENBROCK 2015, S. 58.
[353] Vgl. HABERBERGER/TALARCZYK 2007, S. 369.
[354] Abbildung 9: eigene Darstellung.
[355] Vgl. hierzu ACKERMANN/CLAGES/ROLL 2011, S. 252 ff.

beizumessen. So können die Auswirkungen von Fehlern aufgrund verschiedener Zeitzonen in den Quellsystemen oder der Nutzung von Sommer- und Winterzeit erheblich sein.

Geodaten können auf Karten durch ein Geografisches Informationssystem (GIS) dargestellt werden, die zum einen explorativ durch den Kriminalisten im Rahmen von Visual Analytics ausgewertet werden können. Zum anderen bieten GIS-Anwendungen wie z. B. ESRI ArcGIS[356] oder Uncharted GeoTime[357] die Möglichkeit, mittels Berechnungen Treffpunkte oder besondere Verhaltensweisen von einer Vielzahl von Personen zu erkennen und dann auf einer Karte zur Kontrolle darzustellen. Durch die Kombination von Geodaten und anderen Datenquellen kann der Wert dieser Informationen nochmals deutlich gesteigert werden.[358] So können die Gewohnheiten einer Person aufgehellt werden, wenn die Positionsdaten mit Informationen über die Eigenschaft des Ortes sowie der Dauer und Häufigkeit des Aufenthalts kombiniert werden. Beispielsweise kann der zweimalige Aufenthalt von zwei Stunden pro Woche in einem Gebäude, in dem sich auch ein Sportstudio befindet, den Schluss nahelegen, dass die Person regelmäßig an diesem Ort Sport treibt. Die Informationen, welche besonderen Orte sich an bestimmten Geopositionen befinden, können sowohl in kommerziellen als auch in freien Datenbanken (z. B. OpenStreetMap[359]) abgerufen und automatisiert abgeglichen bzw. zu dem Verfahrensdatenbestand hinzufügt werden. Auch kann die Analyse von Bewegungsspuren mehrerer Personen Hinweise auf mögliche Treffpunkte und Bekanntschaften von Personen geben. Hierfür sind nicht unbedingt telefonische Kontakte zwischen diesen Personen erforderlich.

4.5.4 Statistische Analyse

Eine weitere Analyseform von Big Data ist die statistische Analyse. Hier liegt der Fokus auf der Generierung von Kennzahlen. In der Wirtschaft wird Kennzahl als

> „Zusammenfassung von quantitativen, d.h. in Zahlen ausdrückbaren Informationen für den innerbetrieblichen *(betriebsindividuelle Kennzahlen)* [Hervorhebung im Original; Anm. des Autors] und zwischenbetrieblichen *(Branchen-Kennzahlen)* Vergleich (etwa Betriebsvergleich, Benchmarking) [. . .].“[360]

definiert. In Ermittlungsverfahren ist eine Kennzahl ebenfalls eine in Zahlen ausdrückbare Information, die einen Vergleichswert darstellt. Der Vergleichswert dient

[356] https://www.esri.de (besucht am 05.12.2017).
[357] https://geotime.com (besucht am 05.12.2017).
[358] Vgl. KING 2014, S. 51.
[359] https://www.openstreetmap.org (besucht am 05.12.2017).
[360] Vgl. KRIEGER et al. o. D.

dabei der Verdachtsgenerierung und -verdichtung. Als einfaches Beispiel kann hier die Kontakthäufigkeit zwischen einzelnen Personen benannt werden. Durch Ergänzung weiterer Dimensionen wie z. B. der Festlegung eines Zeitraums (Anrufe pro Tag) und Informationen zum Inhalt des Gesprächs (z. B. durch Schlagwortvergabe) können neue komplexe Informationen für das Ermittlungsverfahren gewonnen werden, die als Grundlage weiterer Ermittlungen dienen.

Die Darstellung der Ergebnisse kann durch Auflistung von Zahlen, aber auch in Form von Diagrammen erfolgen. Danach erfolgt die Interpretation der Ergebnisse durch den Analysten, sodass auch bei dieser Analysemethode Visual Analytics eine wichtige Rolle einnimmt.

Zur Erstellung solcher Kennzahlen können einfache Datenbankenabfragen bis hin zum Einsatz hoch entwickelter Technologien wie z. B. OnLine Analytical Processing-Anwendungen (OLAP)[361] genutzt werden. Die OLAP-Technologie wird vor allem bei umfangreichen Data-Warehouse-Inhalten eingesetzt. Diese erlauben eine Aggregation[362] und Analyse von Daten in Echtzeit. Hierzu werden alle Einzeldaten z. B. für bestimmte Zeitperioden (z. B. Summe aller Transaktionen in Quartalen) oder nach anderen Gruppierungskriterien (z. B. Anzahl der Telefonverbindungen gegliedert nach der Art der Verbindungen wie Sprache, Daten, SMS, etc.) zusammengerechnet. Für kleinere Datenbestände bieten auch verschiedene Softwareprodukte Funktionalitäten zur Erstellung solcher Kennzahlen an. So erlauben z. B. die Tabellenkalkulation Microsoft Excel mit ihrer PivotTabellen-Funktion oder HumanIT InfoZoom mit der Analysengruppen-Funktion umfangreiche Auswertungsmöglichkeiten auf Grundlage verschiedener Datenquellen.

4.5.5 Zusammenfassende Betrachtung und Kernbereichsschutz

Die vorgestellten Analysemethoden haben gezeigt, dass Big Data-Bestände im kriminalistischen Umfeld auf unterschiedliche Weise analysiert werden können. Grundgedanke ist dabei, dass Daten Personen, Sachen und Vorgänge in der realen Welt widerspiegeln. Damit erlaubt die Analyse dieser Daten den kriminalistischen Erkenntnisgewinn an einem Surrogat. Mit Hilfe dieses Abbildes der realen Welt wird versucht, die Position einzelner Personen im Beziehungsgeflecht zu reproduzieren. Durch die fortschreitende Digitalisierung in Wirtschaft und Verwaltung entstehen immer größere Datenbestände, die einzelne Attribute eines Menschen wie z. B. Student, Barzahler, Autobesitzer etc., definieren. Die Kombination von verschiedenen Attributen zu einer Person kann dabei sehr unterschiedlich sein und

[361] Vgl. MERTENS/BODENDORF et al. 2017, S. 52.

[362] „Aggregationen sind Operationen auf Daten, die eine Verdichtung von Daten von einer feineren zu einer gröberen Granularität mittels einer Aggregatfunktion vornehmen. Eine Aggregationsfunktion verdichtet einen Datenbestand bestehend aus n Einzelwerten auf einen einzelnen Wert. Standardaggregationsfunktionen sind SUM(), AVG(), MIN(), MAX() und Count()." FARKISCH 2011, S. 14.

erlaubt unter Umständen die Identifizierung Einzelner. Hier setzt die Mengenanalyse – z. B. als Rasterfahndung – an, indem verschiedene Datenbestände anhand bestimmter Parameter und in Bezug zueinander analysiert werden. Da Personen nicht losgelöst betrachtet werden können vom Ort ihres Aufenthalts – vor, während und nach der Tat – spielt die Analyse der geografischen Daten eine wesentliche Rolle bei der Analyse des Tatgeschehens. Die Daten hierfür ergeben sich entweder direkt aus dem Datenbestand oder können indirekt hinzugefügt werden. Die Kombination einzelner Analysemethoden kann die Informations- und Wissensqualität nochmals steigern. So erlaubt eine Kombination von Netzwerk- und geografischer Analyse die Erweiterung des sozialen Netzwerkes von Tätern und Tätergruppierungen. Die Verbindung von Statistik mit Kommunikationsnetzwerken in Verbindung mit den Zusatzinformationen, die aus Datenabgleichen der Polizei angereichert wurden, kann u. U. zu neuen Tatverdächtigen führen, die aufgrund der isolierten Betrachtung der Ergebnisse zuvor noch nicht aufgefallen sind.

Das Zusammenführen einer Vielzahl von Daten und deren Auswertung birgt jedoch die Gefahr in sich, dass der Kernbereich privater Lebensgestaltung eines Betroffenen berührt wird. Dies beinhaltet nicht nur ein ethisches, sondern auch ein rechtliches Problem. Dieser Kernbereich wurde erstmals 1957 durch das Bundesverfassungsgericht benannt und entzieht dem Staat die Möglichkeit des Zugriffs auf einen unantastbaren Bereich menschlicher Freiheit[363]. Dieser Grundsatz wurde später durch das BVerfG in weiteren Urteilen immer wieder bestätigt und verfeinert, betrifft jedoch immer nur die zugrunde liegende Erhebungsnorm, wie z. B. die Online-Durchsuchung.[364] Auch wenn eine Datenerhebung keinen Kernbereichsbezug aufweist, kann die Kombination und Analyse der daraus gewonnenen Daten diesen dennoch verletzen. So kann z. B. die Auswertung von Orten, an denen sich eine Person aufhält, in Kombination mit Telekommunikationsdaten und Geldbewegungen u. U. Aufschluss über sexuelle Vorlieben einer Person offenlegen. Eine gesetzliche oder gerichtliche Klarstellung der Grenzen der Auswertung im Hinblick auf den Kernbereich privater Lebensgestaltung liegt aktuell nicht vor.

Daneben ist zu beachten, dass solche Ergebnisse immer auf unvollständigen Daten beruhen. So können aus rechtlichen Gründen nicht alle Daten erhoben werden. Außerdem erfolgt die Datenauswahl nach Selektionskriterien, die von den Ermittlern bestimmt sind, auch wenn dabei versucht wird, durch eine systematische Hypothesenbildung eine einspurige Datenerhebung zu vermeiden. Letztendlich handelt es sich bei den Analyseergebnissen immer noch um subjektiv geprägte Hypothesen, die durch weitere Ermittlungen – wie z. B. Vernehmungen – gefestigt werden müssen. Zur Feststellung, wie belastbar die Ergebnisse sind, müssen die bei der Datenerhebung vorgenommene Informationseinstufungen (siehe Ka-

[363] Vgl. BVerfG, Urteil vom 16.01.1957 – 1 BvR 253/56, BVerfGE 6, 32-45.
[364] Vgl. BVerfG, Urteil vom 27.02.2008 – 1 BvR 370/07, BVerfGE 120, 274-350.

pitel 4.8) in geeigneter Form mitgeführt und dargestellt werden. Hierzu müssen die Ergebnisse so dokumentiert werden, dass sie in das Verfahren eingebracht und durch die anderen Verfahrensbeteiligten nachvollzogen werden können. Hierfür dient der nächste Schritt im kriminalistischen Big Data-Analysezyklus.

4.6 Darstellung des Ergebnisses

Im vorangegangenen Prozessschritt wurde aufgezeigt, wie Daten entsprechend kriminalistischer Hypothesen oder explorativ mittels Data Mining und Visual Analytics ausgewertet werden können. Diese Ergebnisse müssen nun in eine für die Nutzer verständliche Darstellungsform überführt werden, da die Ergebnisse u. U. noch für die Zielgruppe unverständlich sind.

Die Ergebnisdarstellung kann entweder in Berichtsform, als Visualisierung oder in einer Kombination aus beiden Darstellungsformen erfolgen. Da die Dokumentation in der Regel als Visualisierung erfolgt,[365] wird die grafische Darstellung der Ergebnisse nachfolgend betrachtet.

Die Notwendigkeit einer Visualisierung ergibt sich aus der Tatsache, dass ein Mensch nur begrenzt Informationen aus schriftlichen Unterlagen entnehmen kann. Durch eine visuelle Darstellung wird das menschliche Denken von der Aufgabe, die ihm vorliegenden Informationen zu strukturieren, entlastet,[366] wodurch die dargestellten Informationen leichter von ihm aufgenommen werden können.

Die Visualisierung der Ergebnisse im Big Data Prozess hat daher zwei Ziele. Zum einen dient sie der Analyse selbst und damit dem Verständnis der Daten sowie der Ergebnisse des Analyseprozesses (Visual Analytics). Andererseits hat die Visualisierung die Aufgabe, die Ergebnisse an eine Person zu kommunizieren, die nicht am Analyseprozess beteiligt war/ist.[367] Die Kommunikation der Informationen soll den Empfänger in die Lage versetzen, die dargestellten Sachverhalte zu erkennen, zu verstehen und zu bewerten und danach sich über die Arbeitsergebnisse mit anderen auszutauschen.[368] Aus den so vermittelten Informationen soll der Empfänger eigenes Wissen generieren, aus welchem er Entscheidungen ableiten kann.[369] Diese Entscheidungen können weitere Ermittlungshandlungen auslösen oder bereits eine Entscheidung eines Staatsanwaltes oder eines Richters sein. Die Visualisierung stellt somit ein wichtiges Instrument dar, relevante Informationen an Entscheidungsträger, wie z. B. Staatsanwalt und Richter, zu transportieren.

Maßgeblich ist dabei, dass die Visualisierung für den Empfänger leicht verständlich und die darin enthalten Informationen leicht verfügbar sind, da nur verstandene

[365] Vgl. ESCH 2009, S. 680.
[366] Vgl. KREMPEL 2005, S. 25 f.
[367] Vgl. SCHUMANN/W. MÜLLER 2000, S. 5.
[368] Vgl. SCHUMANN/W. MÜLLER 2000, S. 5.
[369] Vgl. APEL/BEHME et al. 2015, S. 237.

Informationen das Wissen des Nutzers und somit seine Entscheidungsmöglichkeiten beeinflussen können.[370] Der Aspekt der leichten Verfügbarkeit ist deshalb so wichtig, weil leicht verfügbare Informationen bei Entscheidungsprozessen ein höheres Gewicht haben.[371] Dadurch ist aber auch andererseits eine Beeinflussung des Empfängers durch zielgerichtete oder mangelhafte Darstellung gegeben.

Insbesondere komplizierte und komplexe Sachverhalte sollten über Visualisierungen kommuniziert werden, da sie parallel Informationen liefern, über zusätzliche Ausdrucksformen verfügen und Informationen sehr stark verdichten.[372] Dabei müssen Informationen in der Darstellung effektiv visuell kodiert werden. Durch Ordnung und Zusammenführung mit weiteren Informationen werden diese Informationen besser in konsumierbare Informationseinheiten transkribiert und können leichter vom menschlichen Verstand aufgenommen werden.[373]

Das Erzeugen einer Visualisierung durchläuft selbst einen eigenen Datenaufbereitungsprozess. Dabei werden die Ergebnisdaten der Analyse, die aus dem Blickwinkel der Visualisierung Rohdaten sind, entsprechend der neuen Zieldefinition (Bilderzeugung) vorverarbeitet. Dazu kann es erforderlich sein, weitere für die Visualisierung erforderliche Daten, die für die Analyse nicht erforderlich waren, vom ETL-Prozess-Schritt *Aufbereitung der Daten* (Kapitel 4.4) bis zu diesem Prozessschritt mitzuführen.

Im Schritt des **Filtering** werden irrelevante Daten aus dem Datenbestand entfernt und nur ausgewählte und für das Ziel der Visualisierung notwendige Attribute behalten.[374] Zusätzlich werden in diesem Schritt eventuelle Fehler korrigiert oder von der Visualisierung ausgeschlossen.[375] Die Datenreduktion kann dabei die gesamte darzustellende Datenmenge (Datenbreite), aber auch Umfang und Ausgestaltung der beschreibenden Elemente eines abzubildenden Datensatzes (Datentiefe) betreffen. Eine solche Reduzierung ist schon allein deshalb notwendig, „da sich große Datenmengen auf Grund ihres großen Informationsgehalts selten in einem einzigen Bild veranschaulichen lassen. Das Bild wäre überladen und in der Regel nicht mehr interpretierbar.“[376] Unter Umständen kann es jedoch nötig sein, anstelle einer Datenreduktion die Rohdaten gezielt zu ergänzen, um eine aussagekräftige Visualisierung zu ermöglichen.

Im anschließenden **Mapping**, welches das Kernstück des Visualisierungsprozesses ist, werden die Daten in visuelle Variablen überführt.[377] Bei visuellen Variablen handelt es sich um die Beschreibung der Darstellung (z. B. Farbe, geometri-

[370] Vgl. APEL/BEHME et al. 2015, S. 237.
[371] Vgl. MANGOLD 2008, S. 263 f.
[372] Vgl. APEL/BEHME et al. 2015, S. 246.
[373] Vgl. KREMPEL 2005, S. 25.
[374] Vgl. SCHUMANN/W. MÜLLER 2000, S. 47.
[375] Vgl. SCHUMANN/W. MÜLLER 2000, S. 16.
[376] SCHUMANN/W. MÜLLER 2000, S. 47.
[377] Vgl. SCHUMANN/W. MÜLLER 2000, S. 16.

sche Figur, Größe, Position, etc.) in der Grafik.[378] Die Zuweisung der Darstellungsform zu den Daten hat entscheidenden Einfluss darauf, wie die Ergebnisse vom Betrachter wahrgenommen werden. Dabei unterliegt das Mapping verschiedenen nichttechnischen Einflussfaktoren. Neben Art und Struktur der Daten sowie dem Ziel der Darstellung sind das Vorwissen des Betrachters, dessen visuelle Fähigkeiten (z. B. Farbblindheit) und Vorlieben sowie fachspezifische Konventionen (z. B. spezielle Symbole) wichtige Einflüsse auf die Art und Weise der Darstellung.[379]

Der letzte Schritt im Prozess ist das **Rendering**, welches die grafische Darstellung der Daten generiert.[380]

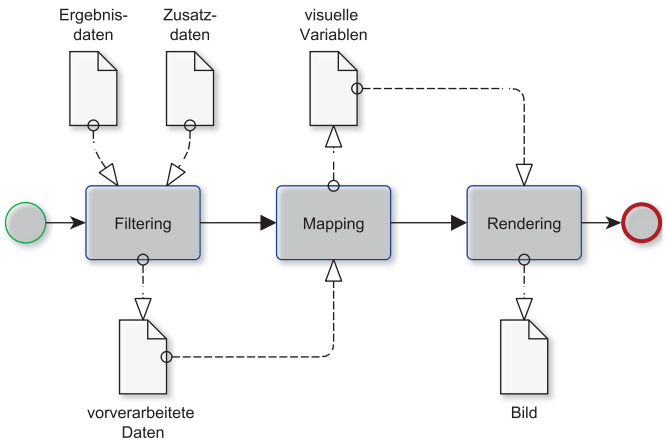

Abbildung 10: Visualisierungsprozess[381]

4.7 Dokumentation des Prozesses

Nachdem die Ergebnisse dargestellt und kommuniziert wurden, ist ein weiterer Prozessschritt die Dokumentation des gesamten Auswertungsprozesses. Die Dokumentation soll hierbei als „[. . .] Beschreibung eines Sachverhalts, der in sich abgeschlossen ist, ein gewisses Maß an Vollständigkeit erreicht hat und zur (dauernden) Aufbewahrung bestimmt ist [. . .]“[382] verstanden werden. Dabei beinhaltet der Sachverhalt das Zustandekommen des Auswertungsergebnisses. Dieser ist in Anlehnung an die Definition von Wilhelm Gaus vollständig zu beschreiben und soll dann zur Aufbewahrung zur Ermittlungsakte genommen werden.

Clages erachtet eine solche Dokumentation im Rahmen eines Ermittlungsverfahrens für zwingend geboten. Nach seiner Auffassung muss eine Dokumentation aller Ermittlungshandlungen umfassend und lückenlos sein.[383] Dieser Sichtweise

[378] Vgl. Bertin 1982, S. 186.
[379] Vgl. Schumann/W. Müller 2000, S. 8 f.
[380] Vgl. Schumann/W. Müller 2000, S. 16.
[381] Abbildung 10: eigene Darstellung in Anlehnung an Schumann/W. Müller 2000, S. 15.
[382] Gaus 2003, S. 33.
[383] Vgl. Clages 2017a, S. 64.

ist vor allem im Bereich der Auswertung von Big Data zuzustimmen, da im Laufe des Prozesses eine Vielzahl von Eingriffen erfolgt. Ein besonderes Augenmerk muss dabei auf die Sequenzstellen[384] im Prozess gelenkt werden, an denen erstens ein Sachbearbeiter durch seine Entscheidungen den Prozess gelenkt hat und zweitens die Daten von ihm bearbeitet, also gefiltert, ergänzt, gelöscht oder auf andere Art geändert wurden. Ohne eine solche Dokumentation können diese Bearbeitungen und Veränderungen später nicht mehr rekonstruiert werden, da lediglich die umfangreichen Ausgangsdaten und das Auswertungsergebnis für eine Prüfung durch die Staatsanwaltschaft, Richter und Verteidiger zur Verfügung stehen. Eine solche Prüfung muss aber jederzeit möglich sein, da sonst ein rechtsstaatliches Verfahren nicht möglich ist. Diese Verpflichtung besteht nicht nur für manuell durchgeführte Erhebungs- und Auswertungsmethoden, sondern gerade auch für automatisierte Auswertungen,[385] die, wie die bisherigen Ausführen gezeigt haben, durch manuelle Zwischenschritte unterbrochen sind.

Eine besondere Herausforderung stellt dabei die verständliche Darstellung des Verlaufs der Big Data-Auswertung für Außenstehende dar. So muss die Dokumentation geeignet sein, anderen Personen den Hergang des Auswertungsprozesses verständlich zu machen. Daher sind auch hier Beschreibungs- und Visualisierungsmethoden erforderlich, um den Prozess der Big Data-Auswertung an Personen zu kommunizieren, die selbst an diesem Prozess nicht beteiligt gewesen sind.

Grundsätzlich erfolgt die Dokumentation von Ermittlungsergebnissen und deren Gewinnung in schriftlicher bzw. bebilderter Form, da nur die Informationen von Richtern und Staatsanwälten für weitere Entscheidungen zählen, die in dieser Form vorliegen.[386] Dies ist auch darin begründet, dass eine Dokumentation in Aktenvermerken für das Ermittlungsverfahren als Beweistatsache dient.[387]

Auch wenn vor Gericht das Prinzip der Unmittelbarkeit gilt, führt nur eine schriftliche und visuelle Fixierung des gesamten Prozesses in der Ermittlungsakte dazu, dass diese Informationen durch alle anderen Beteiligten im Strafprozess auch außerhalb der Ermittlungsbehörde nachvollzogen werden können. Zudem sind die Behörden zur Erstellung von Akten zur vollständigen Dokumentation ihres Handelns verpflichtet.[388]

Neben der Dokumentation des Prozesses in schriftlicher Form (Bericht) können wieder Visualisierungstechniken verwendet werden. Diese können z. B. bei der Darstellung der Data Lineage (Datenabstammungsbeschreibung; siehe Abbildung 11) zum Einsatz komme und beschreiben anhand von Metadaten den Ver-

[384] Sequenzstellen sind „[...][d]ie Stellen des Handlungsablaufes, an denen sich durch Regeln sinnlogische Handlungsmöglichkeiten eröffnen[...].“; BRISACH 1992, S. 180.
[385] Vgl. KÖRFFER 2014, S. 150.
[386] Vgl. JAEGER 2005, S. 420.
[387] Vgl. WEIHMANN/VRIES 2014, S. 390.
[388] Vgl. BVerwG, Beschluss vom 16.03.1988 – 1 B 153/87.

änderungsprozess der Daten vom Rohdatum bis zum Auswertungsergebnis.[389] Diese Dokumentation sollte neben dem Datenfluss auch das Regelwerk für die

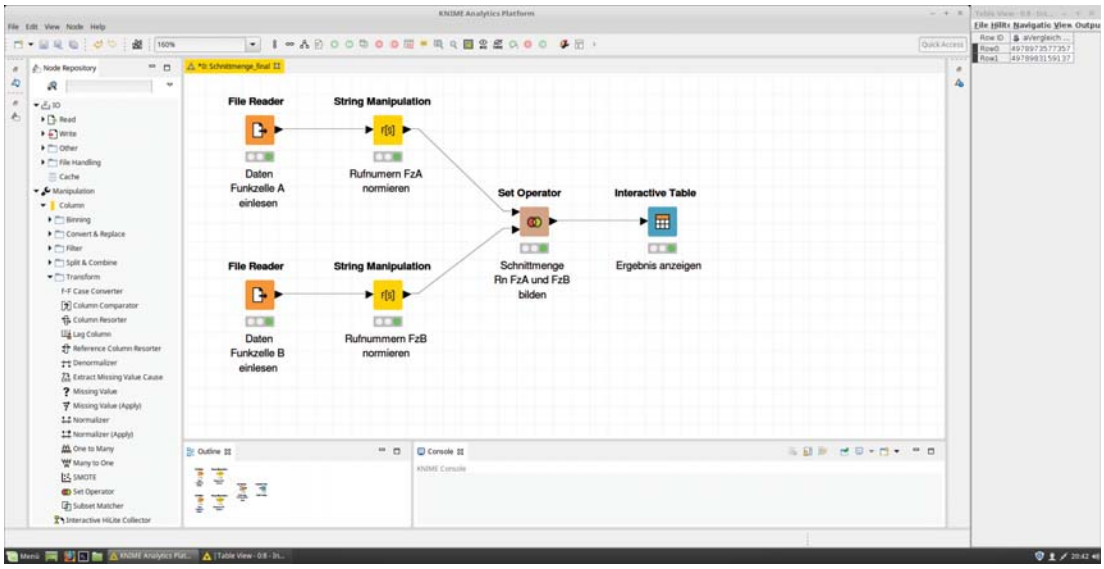

Abbildung 11: Beispiel einer Data Lineage zur Dokumentation einer Funkzellenauswertung[390]

Transformation der Daten enthalten. Farkisch[391] beschreibt die Notwendigkeit eines solchen Regelwerkes für die Erstellung eines Data-Warehouse, um zu einem späteren Zeitpunkt die Ergebnisse und deren Zustandekommen validieren zu können. Diese Notwendigkeit kann ebenfalls im Prozess zur Datenaufbereitung von Big Data im Ermittlungsverfahren anerkannt werden.

Zurückblickend auf die bisherigen Ausführungen sollte eine Dokumentation mindestens folgende Punkte enthalten:

1. Umstände der Datenerhebung
2. Quellenbewertung der Ausgangsdaten
3. Ergebnisse des Data Profilings
4. Wesentliche Schritte in der Datenaufbereitung, vor allem
    a) Regeln bei der Datenextraktion
    b) Datenveränderungen
    c) Datenbereinigungen
5. Wesentliche Schritte der Datenanalyse
6. Wesentliche Schritte der Ergebnisdarstellung, vor allem
    a) Beschreibung der visuellen Variablen
    b) Datenveränderungen
    c) Datenbereinigungen

[389] Vgl. BITCOM 2014, S. 98.
[390] Abbildung 11: eigene Darstellung.
[391] Vgl. FARKISCH 2011, S. 63.

4.8 Evaluation des Ergebnisses und des Prozesses

Im letzten Schritt des kriminalistischen Big Data-Auswertungszyklus erfolgt eine Evaluierung der Ergebnisse mit Blick auf die zu Beginn aufgestellten Hypothesen und Untersuchungsfragen. Kern der Betrachtung ist hierbei, ob die Daten und die darauf aufbauende Analyse geeignet waren, die Untersuchungsfragen zu beantworten. Im negativen Fall müssen erneut Daten erhoben oder die Analyse mit veränderten Parametern erneut durchgeführt werden. Also beginnt der Analysekreislauf von Neuem, bis keine ungeklärten Untersuchungsfragen mehr offen sind.[392]

4.9 Zusammenfassung

Die Ausführungen in diesem Kapitel haben gezeigt, dass es sich beim kriminalistischen Big Data-Auswertungszyklus um eine komplexe Abfolge von acht aufeinander aufbauenden Einzelschritten handelt, die ineinander übergehen und voneinander abhängig sind. Ausgehend von einer Hypothesenbildung, die auf Basis einer Fallanalyse die noch offenen Informationslücken mit Annahmen schließt, werden mehrere Untersuchungsfragen formuliert, die schließlich in eine Maßnahmenplanung münden (Abschnitt 4.1). Bei der Maßnahmenplanung sind die rechtlich zulässigen Erhebungsmethoden zu wählen, bei denen auch angenommen werden kann, dass die erhobenen Daten auch tatsächlich ausgewertet werden können. Hierbei spielen neben fachlichen Einschränkungen wie z. B. erwartete Datenqualität vorrangig materielle und personelle Ressourcen eine wesentliche Rolle. Nach der Datenerhebung muss die Qualität der Datenquellen bewertet werden (Abschnitt 4.2). Diese Einstufung ist für die Bewertung der Zulässigkeit von Auswertungsergebnissen erforderlich, da diese zum Teil auf unsicheren Daten beruhen könnten und somit weitere verifizierende Maßnahmen benötigen. Der Datenerhebung und Quellenbewertung folgt das Data Profiling, wodurch Metadaten erzeugt werden, die Aussagen über die Qualität der Daten treffen (Abschnitt 4.3). Diese Metadaten werden sowohl im anschließenden ETL-Prozess, in der Prozessdokumentation und in der Prozess- und Ergebnisevaluation benötigt. Zudem sind diese Metadaten für eine erneute Maßnahmenplanung erforderlich. Data Profiling hat damit entscheidenden Einfluss auf die Ausgestaltung der folgenden Handlungsschritte im kBDA. Der technische Kernprozess des kBDA ist die Aufbereitung der Daten (Abschnitt 4.4). Der Umfang des ETL-Prozesses richtet sich nach dem Ziel der Auswertung, der Datenqualität und der Analyseanwendung. In diesem Prozessschritt werden die Daten z. B. aus Datenbanken oder semi- bzw. unstrukturierten Datenquellen extrahiert und anschließend in einem Transformationsschritt vereinheitlicht. Die Aufbereitung ist mit dem Laden der Daten in ein für die Analyseanwendung nutzbares

[392] Vgl. WALDER/HANSJAKOB 2016, S. 94.

Datenformat abgeschlossen. Bei diesem Schritt werden neben den Analysedaten noch weitere Metadaten bzgl. der Art und Weise des ETL-Prozesses generiert, die für die Prozessdokumentation erforderlich sind. Zudem müssen im Hinblick auf die an die Analyse anschließende Visualisierung unter Umständen bereits Visualisierungsrohdaten erzeugt und als Metadaten den Analysedaten mitgegeben werden. Der fachliche Kernprozess ist die sich an die Aufbereitung anschließende Analyse (Abschnitt 4.5). Diese richtet sich grundlegend nach den gestellten Ermittlungsfragen. Daher ist eine enge Verzahnung von Ermittlungen und technischer Analyse unabdingbar. Nach Durchführung der Analyse wird das Ergebnis im Visualisierungprozess vornehmlich grafisch dargestellt (Abschnitt 4.6). Hierfür sind neben den Ergebnisdaten unter Umständen auch weitere Metadaten erforderlich, die im ETL-Prozess generiert wurden. Für die Nachvollziehbarkeit des konkreten Auswertungsprozesses ist die Dokumentation aller Handlungen erforderlich (Abschnitt 4.7). In diese Dokumentation fließen die Umstände der Datenerhebung, des Ergebnis des Data Profilings, der Umfang der Datenaufbereitung, die Durchführung der Datenanalyse und der Visualisierung ein. Nur durch eine geeignete Dokumentation kann der gesamte Prozess einer rechtsstaatlichen Prüfung zugänglich gemacht werden. Nach Erstellung der Dokumentation erfolgt abschließend eine Evaluation der Ergebnisse und des Prozesses unter Berücksichtigung der Ziele der Auswertung (Abschnitt 4.8). Sollte es danach erforderlich sein, sind einzelne Teile erneut zu durchlaufen oder der Prozess abermals zu beginnen.

5 Gesamtfazit

In der hier vorgelegten Arbeit sollte dargestellt werden, wie Big Data in Ermittlungsverfahren ausgewertet werden kann. Um sich dem Thema zu nähern, war es erforderlich, den Begriff Big Data zu bestimmen. Dabei konnte festgestellt werden, dass eine Bestimmung des Begriffs als große Datenmenge diesem nicht gerecht wird. So spielen neben dem Datenvolumen auch die Datenvielfalt, die Geschwindigkeit der Datenentstehung und Datenverarbeitung, die Glaubwürdigkeit der Ergebnisse und die Werthaltigkeit der Datenauswertung eine Rolle bei der Begriffsbestimmung. Diese Eigenschaften beeinflussen sich gegenseitig. Aufgrund eines großen Datenvolumens muss eine Datenauswertung schneller erfolgen, um Ergebnisse in der gleichen Zeit vorliegen zu haben. Diese Verarbeitungsgeschwindigkeit wird dabei zusätzlich beeinflusst durch die Vielfalt der Datenquellen und der damit einhergehenden Vielfalt der Datenabbildung, die eine umfangreichere Bearbeitung der Daten zur Folge hat. Die Heterogenität der Datenquellen wirkt sich zudem auf die Glaubwürdigkeit der Ergebnisse aus, wodurch die Werthaltigkeit, also die Nützlichkeit der Ergebnisse, nachteilig beeinflusst werden kann. Diese Eigenschaften von

Big Data und deren Abhängigkeiten voneinander können auch bei Daten, die im Rahmen von Ermittlungsverfahren erhoben werden, festgestellt werden. Dadurch ist das Thema Big Data kein exklusives Thema der Wirtschaft, sondern auch ein Thema der Strafverfolgung. In der gesamten Diskussion wurde deutlich, dass die bloße Datenmenge aus fachlicher Sicht nicht das entscheidende Problem in der Bearbeitung von Big Data ist. Dieser Einflussfaktor soll und kann nicht vernachlässigt werden, da die Datenmengen Auswirkungen auf personelle sowie materielle Ressourcen haben. Vielmehr wirkt sich die Big Data-Eigenschaft Vielfältigkeit (Variety) auf viele Prozessschritte bei der Bearbeitung und Verwertung dieser Datenbestände aus. Aufgrund der Vielfältigkeit von Daten sind umfangreiche Maßnahmen zur Qualitätssicherung und zur Vereinheitlichung der Daten erforderlich, die ca. 70-80 % der Arbeitszeit in Anspruch nehmen. Der Vielfältigkeit, aber auch der Datenmenge ist es geschuldet, dass es zum Erkenntnisgewinn und zur Ergebnisweitergabe erforderlich ist, komplexe Daten- und Informationsstrukturen zu vereinfachen. Dieser Vereinfachungsprozess erfolgt dabei auf Grundlage individueller Selektionsmechanismen, die verschiedenen Verzerrungen unterliegen können.

Ebenso schwerwiegend ist das Kriterium der Glaubwürdigkeit (Veracity) der Auswertungsergebnisse von Big Data. Wie festgestellt wurde, ist diese abhängig von der Datenqualität und von der Pluralität der Datenquellen. Da ein Ermittlungsverfahren mit schwerwiegenden Grundrechtseingriffen verbunden sein kann, muss das Kriterium Glaubwürdigkeit besonders betrachtet werden. Dieser Umstand verlangt, dass die fachliche und technische Qualität sowie die Herkunft der in einem Ermittlungsverfahren verarbeiteten Daten jederzeit nachvollziehbar sein muss, damit die Zuverlässigkeit der Auswertungsergebnisse zu jeder Zeit feststellbar ist. Hierfür sind die Daten, das Ergebnis und die Datenverarbeitungsprozesse nachvollziehbar zu dokumentieren.

Neben der Begriffsbestimmung von Big Data war für die eingangs aufgestellte Frage eine Abgrenzung der Begriffe Daten, Informationen und Wissen erforderlich. Hier konnte festgestellt werden, dass Daten nicht mit Informationen gleichgesetzt werden können. Daten, die grundsätzlich nur eine Reihe von Zeichen darstellen, sind erst dann Informationen, wenn diese eine entsprechende Bedeutung erhalten. Erst mit Bestimmung des Kontextes durch Metadaten wird aus einem Datum eine Information. Der entscheidende Schritt für die Verwertung von Big Data ist dann die Transformation von Information zu Wissen. Diese Transformation ist – im Gegensatz zum Schritt von Daten zu Informationen – ein subjektiver Vorgang. Der Nutzer der Informationen bettet dabei die neuen Informationen in seine bestehenden Wissensstrukturen ein und generiert in seinem Kopf neues Wissen. Aufgrund der Eigenschaften von Big Data und deren Verarbeitung im Auswertungsprozess kommt der Unterstützung des Nutzers für das Verständnis der Ergebnisse eine besondere Bedeutung zu. Es genügt demnach nicht, nur die Daten und Informationen

aufzubereiten und zu analysieren, es ist vielmehr erforderlich, die Ergebnisse so darzustellen, dass sie dem Empfänger auch verständlich sind. Art und Umfang dieser Darstellung richtet sich dabei an dem konkreten Nutzer der Ergebnisse aus. Bei der Erstellung sind daher neben dem Ziel der Auswertung auch individuellen Kenntnisse und Fähigkeiten des Nutzers zu berücksichtigen.

Da in dieser Arbeit Big Data in Ermittlungsverfahren thematisiert wurde, sind Rahmenbedingungen zu beachten, die sich aus dem Rechtsstaatprinzip ergeben. So hat sich gezeigt, dass bei der Auswertung von Big Data in Ermittlungsverfahren grundsätzlich personenbezogene Daten verarbeitet werden, da die Herstellung des Bezugs von Datum und der Person des Täters das festgelegte Ziel der Ermittlungen ist. Aus diesem Umstand ergeben sich Zwänge aus den Datenschutzgesetzten wie z. B. Zweckbindung, Datensparsamkeit und Betroffenenrechte. Insbesondere die Zweckbindung und die Datensparsamkeit stehen dabei mit der Erhebung und Auswertung von Big Data in einem Spannungsverhältnis. So werden für die Auswertung von Big Data-Datenbeständen Daten aus einer Vielzahl von verschiedenen Datenquellen verarbeitet, deren Erhebung ursprünglich zu einem anderen Zweck erfolgte. Nur mittels Zweckänderungen können die Daten umgewidmet werden, was im Ermittlungsverfahren unter Beachtung der entsprechenden rechtlichen Schranken möglich ist. Ebenso steht Datensparsamkeit dem Big Data-Charakteristikum Datenmenge gegenüber. Grundsätzlich sind Polizei und Staatsanwaltschaft verpflichtet, alle Maßnahmen zu treffen, die eine Verdunklung der Straftat verhindern und auch entlastend zu ermitteln. Das betrifft auch die Erhebung aller für das Strafverfahren erforderlichen Daten. Welche Daten für das Verfahren erforderlich sind, lässt sich jedoch selten am Anfang des Ermittlungsverfahrens vorhersehen. Dennoch muss in Zukunft Datensparsamkeit sowohl aus datenschutzrechtlicher Sicht aber auch wegen begrenzter personeller Ressourcen besonders beachtet werden. Hierbei kann eine klare Fokussierung auf beantwortbare Fragestellungen für die Big Data-Auswertung im Ermittlungsverfahren helfen.

Um Daten analysieren zu können, stehen eine Vielzahl von verschiedenen Techniken, die in verschiedenen Auswertungsprozessen eingebettet sind, zur Verfügung. So konnte aufgezeigt werden, dass die in der Wirtschaft genutzten Analysetechniken wie Data Mining, Text Mining, Web Mining sowie Visual Analytics zur Bearbeitung von strafrechtlich relevanten Sachverhalten genutzt werden können.

Nachdem festgestellt werden konnte, dass Big Data auch in Ermittlungsverfahren eine Rolle spielen kann und die in der Wirtschaft genutzten Techniken grundsätzlich geeignet sind, diese auch für kriminalpolizeiliche Daten zu verwenden, war zu prüfen, wie ein solcher Prozess grundsätzlich aussehen kann.

Ausgehend von verschiedenen Auswertungszyklen und -prozessen aus der Wirtschaft und der Kriminalistik wurde in der vorliegenden Arbeit ein neuer Auswertungszyklus entwickelt. Diese Neuentwicklung war erforderlich, weil die bestehen-

den und dargestellten Prozesse sich entweder auf die Big Data-Auswertung im wirtschaftlichen Umfeld bezogen oder nur die Auswertung von geringen Datenmengen im kriminalistischen Umfeld zum Inhalt hatten. Eine einfache Verschmelzung dieser Prozesse genügte nicht, um die Anforderungen von Big Data und Ermittlungsverfahren gleichermaßen zu erfüllen. Vielmehr sind weitere Aspekte zu beachten, die sich einerseits aus potenziellen Gefahren für die Grundrechte der Betroffenen im Rahmen eines Ermittlungsverfahrens und andererseits aus den speziellen Charakteristika von Big Data ergeben. So kann in diesem Zusammenhang nicht nur die Analyse der Daten zur Generierung der Ergebnisse erklärtes Ziel des Auswertungsprozesses sein. Darüber hinaus müssen besondere Rahmenbedingungen beachtet werden. So muss u. a. der Auswertungsprozess unvoreingenommen erfolgen. Grund hierfür sind verschiedene psychologische Aspekte, die zu einer Verzerrung in der Informationswahrnehmung, -verarbeitung und -nutzung führen können. Um solche Verzerrungen möglichst zu vermeiden oder zumindest zu verringern, muss eine Objektivierung des Auswertungsprozesses kultiviert werden. Zudem sind besondere Anforderungen an die Nachvollziehbarkeit der Auswertung zu beachten, damit alle Beteiligten des Strafverfahrens wie andere Ermittler, Staatsanwaltschaft, Richter und Verteidiger die Durchführung der Auswertung rekonstruieren und auf mögliche Fehler prüfen können. Diese Anforderung ergibt sich aus dem Recht auf ein faires Verfahren. Die Schwierigkeit besteht darin, dass der Auswertungsprozess sehr komplex sein kann und ein Verständnis und teilweise fundiertes Wissen bezüglich der Auswertungsmethoden verlangt. Daher stellt die Dokumentation des Prozesses eine besondere Herausforderung an den Sachbearbeiter dar. Die Dokumentation soll dem Empfänger der Ergebnisse das notwendige Hintergrundwissen liefern, um schließlich die Verfahrensinformationen zu Verfahrenswissen transformieren zu können.

Im Ergebnis ergeben sich zwei wesentliche Anforderungen an einen kriminalistischen Auswertungsprozess von Big Data. Das ist zum einen die Unvoreingenommenheit und Objektivität bei der Bestimmung und Erreichung der Auswertungsziele und zum anderen eine Transparenz des gesamten Auswertungsprozesses. Diese Anforderungen spiegeln sich in den acht Punkten des hier entwickelten kriminalistischen Big Data-Analysezyklus‘ wider:

1. Definition der Ziele und Planung der Maßnahmen
2. Erhebung der Daten und Bewertung der Quellen
3. Data Profiling
4. Aufbereitung der Daten
5. Analyse der Daten
6. Darstellung des Ergebnisses
7. Dokumentation des Prozesses
8. Evaluation des Ergebnisses und des Prozesses

So dient eine Hypothesenbildung im ersten Schritt der Objektivierung der Auswertungsprozesse, in dem durch die Bildung von variantenhaften Annahmen über Täter und Tatabläufen eine einseitige Datenerhebung aufgrund individueller Verzerrungsfaktoren verhindert wird. Ebenso unterstützt die Quellenbewertung im zweiten Schritt die objektive Betrachtung der Informationen, indem die Quelle der Informationen neutral bewertet wird und diese Metainformationen – die Einstufung ob es sich um sichere oder unsichere Informationen handelt – weiter mitgeführt werden. Diese Vorgehensweise erlaubt im späteren Verlauf die Beurteilung, ob ein Auswertungsergebnis auf sicheren Informationen beruht. Diese Einstufung ist vor allem aufgrund des Big Data-Charakteristikums Glaubwürdigkeit von großer Bedeutung. Zur Bestimmung der Glaubwürdigkeit der Daten dient auch der Prozessschritt *Data Profiling*, da hier die Datenqualität geprüft wird.

Der fachliche Kernprozess ist die Analyse der Daten, die durch den Schritt 4 – Aufbereitung der Daten – vorbereitet wird. Im Rahmen dieser Arbeit konnten exemplarisch vier Analysemethoden dargestellt werden. Diese Darstellung ist aufgrund des begrenzten Umfangs der Arbeit nicht abschließend. So konnten z. B. Methoden aus der Computerforensik oder der Internetauswertung nicht einbezogen werden. Aus demselben Grund wurde auf eine detaillierte Darstellung dieser Analysemethoden verzichtet.

Die Darstellung dieser Methoden zeigt die Möglichkeiten der Wertschöpfung für ein Strafverfahren. Ausgehend von der Annahme, dass der Mensch als soziales Wesen in soziale Strukturen eingebettet ist, erlaubt die Soziale Netzwerkanalyse eine Hypothese zu Stellungen einzelner Personen in diesem Netz anhand von Kennzahlen zu erstellen. Verschiedene Studien stützen mit ihren Ergebnissen den Ansatz, dass durch Analysen von Kommunikationsverhalten die tatsächlich bestehenden sozialen Kontakte aufgedeckt werden können. Diese datengetriebene Herangehensweise stellt eine Data Mining-Methode dar, deren Ergebnisse im konkreten Ermittlungsverfahren verifiziert werden müssen. Eine solche Abklärung ist aufgrund verschiedener Verzerrungsfaktoren wie die fokussierte Datenerhebung zu einzelnen Personen und der durch den Ermittler konstruierten Relevanzmerkmale erforderlich.

Ein weiterer Ansatzpunkt für Datenanalysen in Ermittlungsverfahren ist die Tatsache, dass strafbare Handlungen immer einen Ort verlangen, an dem der Täter vor, während und nach der Tat gehandelt hat. In der Arbeit konnte gezeigt werden, dass es Datenbestände gibt, die direkte Geodaten enthalten. Zudem können unter Umständen auch Geopositionen indirekt einer Handlung und somit einer Person zugeordnet werden. Anhand dieser Daten lassen sich verschiedene Fragestellungen wie Aufenthalte an tatrelevanten Orten oder Treffen zwischen Mittätern beantworten. Hierfür sind nicht nur Ortsangaben erforderlich, sondern auch die Angaben zum Zeitpunkt, an dem sich der Betroffene an dem jeweiligen Ort befunden hat.

Auch hier kann zum einen ein datengetriebener Ansatz verfolgt werden, wenn aufgrund der Analyse Hypothesen zum Aufenthalt von Personen aufgestellt werden. In diesem Fall wird von Spatial Data Mining gesprochen. Zum anderen können die Geodaten auch auf einer Karte dargestellt und um Rahmen von Visual Analytics bewertet werden.

Ein hypothesengeleiteter Ansatz ergibt sich aus der Mengenanalyse. Hierbei werden entsprechend der vorab in Schritt 1 des kriminalistischen Big Data-Auswertezyklus‘ erstellten Hypothesen die erhobenen Datenbestände auf eine bearbeitbare Datenmenge reduziert. Dabei werden verschiedene Datenbestände in Beziehung gesetzt und damit Schnitt-, Vereinigungs- oder Ausschlussmengen gebildet. Diese Methode wird bei verschiedenen Maßnahmen wie der Funkzellenauswertung, der Rasterfahndung oder bei Datenabgleichen angewendet.

Eine Steigerung des Informationsgehaltes der Auswertung ergibt sich aus einer Kombination der verschiedenen vorgestellten Methoden. So konnte aufgezeigt werden, dass die Kombination von Sozialer Netzwerkanalyse, Geoauswertung und statistischer Auswertung den Informationsgehalt von erhobenen Daten nochmals steigern kann. Hier ist jedoch zu beachten, dass durch eine vollumfängliche Auswertung von Daten der Kernbereich eines Betroffenen verletzt werden kann. Hieraus ergibt sich auch das ethische Dilemma der Big Data-Auswertung, das auch bei der Auswertung innerhalb eins Ermittlungsverfahrens vorliegt. Konkrete gesetzliche Regelungen zur Begrenzung der Auswertung im Hinblick auf den Kernbereich wurden noch nicht geschaffen, sodass auch die Datenverarbeitungsmaßnahmen der Analyse verfassungskonform auszulegen sind.

Nach Durchführung der Analyse erfolgt die Visualisierung der Analyseergebnisse. Die Ausführungen haben gezeigt, dass auch diese verschiedenen psychologischen Verzerrungen unterliegen können. Das bedeutet für die Visualisierung, dass sie zum einen alle Informationen für eine Interpretation der Ergebnisse enthalten muss und dabei zum anderen durch die Art der Darstellung den Betrachter nicht manipuliert. Diese Gefahr besteht vor allem, da zur Visualisierung die Informationstiefe und -breite reduziert werden muss, um eine mental erfassbare Darstellung überhaupt zu ermöglichen. Die für die Visualisierung notwendigen Daten werden nicht nur im Analyseprozess gewonnen, sondern entstehen ebenfalls im Datenaufbereitungsprozess und müssen, wie die Ergebnisse der Quellenbewertung, im Visualisierungsprozess einbezogen werden und sich in der Darstellung der Ergebnisse wiederfinden.

Bevor der gesamte Prozess anhand der Ergebnisse und durch Rückschau auf den gesamten Auswertungsprozess evaluiert wird, muss der Prozess dokumentiert werden. Bei der Diskussion der verschiedenen Einflüsse auf die Auswertung und die Nutzung der Ergebnisse als Spur oder Beweis wurde deutlich, dass eine solche Dokumentation zur Wahrung rechtstaatlicher Prinzipen unumgänglich

ist. Die Dokumentation muss insbesondere die Umstände der Datenerhebung, die Ergebnisse der Datenqualitätsprüfung, den Datenaufbereitungs- und Datenanalyseprozess sowie die Verfahrensweise bei der Ergebnisdarstellung umfassen. Die Erörterungen hierzu haben gezeigt, dass sich nur so die Beteiligten des Strafverfahrens zu jeder Zeit ein Bild über die Entstehung der entsprechenden Spur bzw. des Beweises machen können. Dementsprechend ist die Dokumentation auch so zu erstellen, dass der Grund und die Art und Weise wesentlicher Datenveränderungen aus der Dokumentation hervorgeht.

Der hier vorgestellte Prozess bildet die Möglichkeit der Big Data-Auswertung im Rahmen einer Arbeitsteilung zwischen Ermittler und Analyst ab. Zudem beschreibt dieser eine starke Standardisierung von Prozessen, die sich zunächst einmal in der Praxis etablieren müssen. Daher muss der Prozess zu einem späteren Zeitpunkt durch weitere kriminalistische Forschungen evaluiert und an eventuell neue technische, rechtliche, politische und organisatorische Bedingungen angepasst werden. Eine solche Forschung wäre im Rahmen einer neu zu etablierenden Kriminalinformatik denkbar.

Die Big Data-Auswertung wird kein vorübergehendes Phänomen sein, sodass sich die Ermittlungsbehörden dementsprechend aufstellen müssen. So haben die Ausführungen gezeigt, dass neben dem kriminalistischen Sachverstand bei der Hypothesenbildung und Maßnahmenplanung auch umfangreiches Wissen im Bereich der Informatik erforderlich ist, welches sich vom klassischen Berufsbild des Kriminalisten unterscheidet. Demnach sind zum einen für diese Tätigkeit Spezialisten (Kriminalinformatiker) erforderlich und zum anderen muss ein Ermittler befähigt werden, die Möglichkeiten der Big Data-Auswertung zu erkennen und unter Zuhilfenahme von Spezialisten ergebnisorientiert, objektiv und rechtsstaatlich zu nutzen.

Anlage A

```
import re
import os

re_mail = r'[A-Z0-9a-z._+\-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,64}'
fobj_out = open('NER-EMail.csv','w')

for filename in os.listdir('Rohdaten'):
        fobj_in = open('Rohdaten/' + filename, 'r')
        for line in fobj_in:
                inhalt = line.rstrip()
                mails = re.findall(re_mail, inhalt)
                for email in mails:
                        fobj_out.write(filename + ';' + email + '\r')
        fobj_in.close()

fobj_out.close()
```

Listing 1: Quellcode in Python 3 zur Extraktion von E-Mailadressen aus Textdateien (Quelle: eigene Darstellung)

Anlage B

Anlage B: Synopse der in Kapitel 3 diskutierten Auswertungszyklen

Walter	**Büchler et al.**		**Intelligence Cycle**		**CRISP-DM**	
1 Verdacht	—		—		—	
2 Daten analysieren	6	Informationsanalyse	4	Analysis and Production	4	Modeling
	7	Informationsbewertung/ Schlussfolgerung				
	8	Ergebnisdarstellung				
	9	Ergebnisweitergabe				
—	10	Ergebnisumsetzung	—		—	
—	11	Evaluation und Rückkopplung	5	Dissemination and Integration	5	Evaluation
3 Hypothesen bilden	1	Zielbildung	1	Planing and Direction	1	Business Understanding
4 Programm bestimmen	—				—	
5 Fehlende Daten beschaffen	2	Informationssuche/-sammlung	2	Collection	2	Data Understanding
	3	Informationsaufnahme			—	
—	4	Informationsordnung	3	Processing and Exploitation	3	Data Preparation
—	5	Informationsspeicherung				
—	—		—		6	Deployment

Anlage C

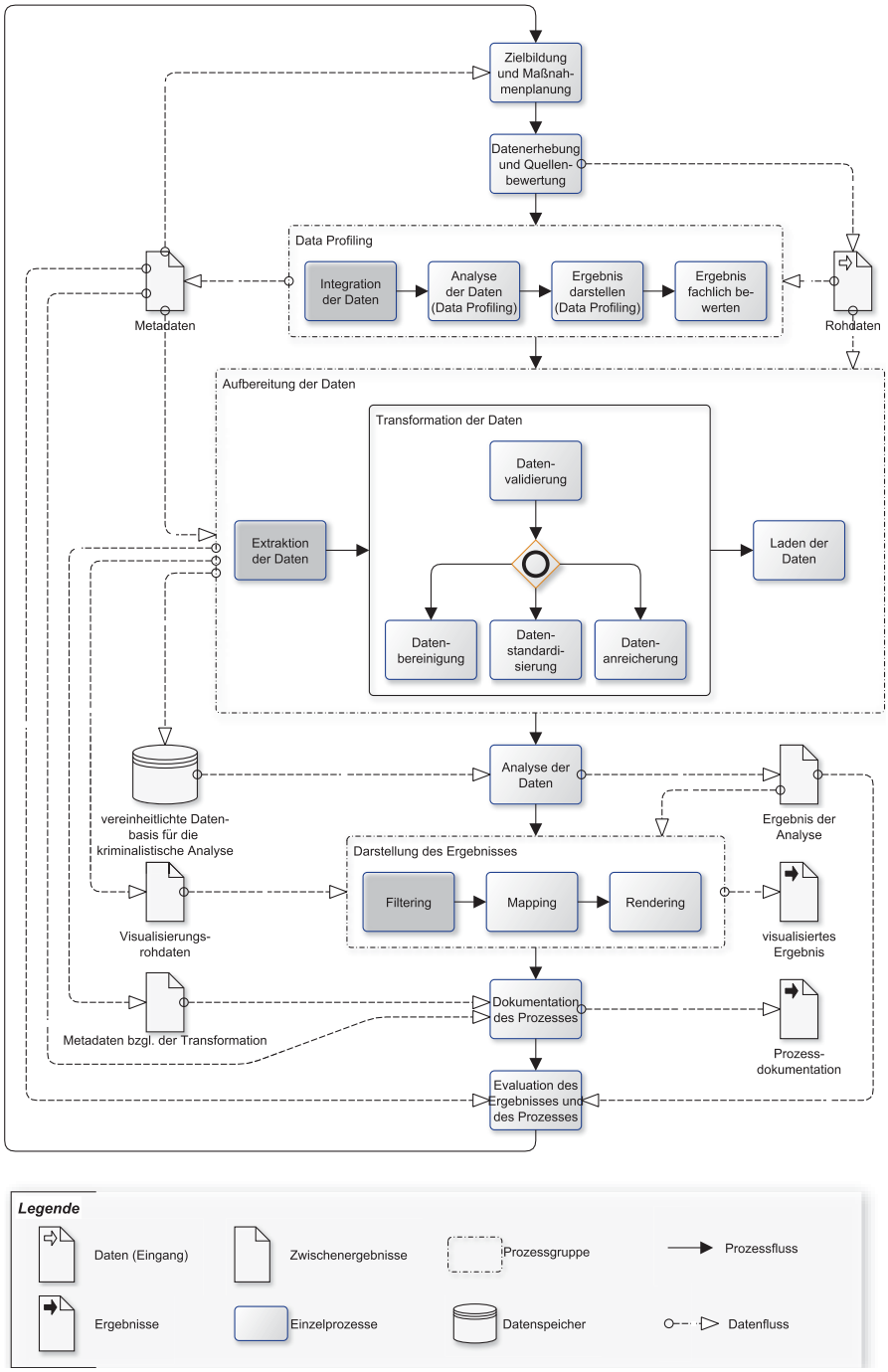

Abbildung 12: detaillierte Abbildung des kBDA[393]

[393] Abbildung 12: eigene Darstellung.

Anlage D

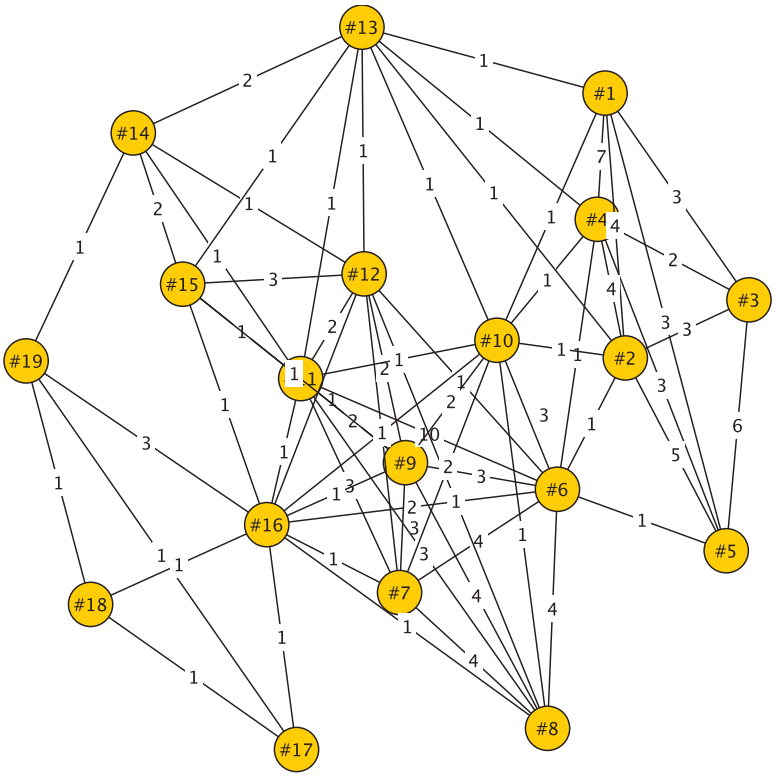

Abbildung 13: gesamtes Netzwerk (ungewichtet)

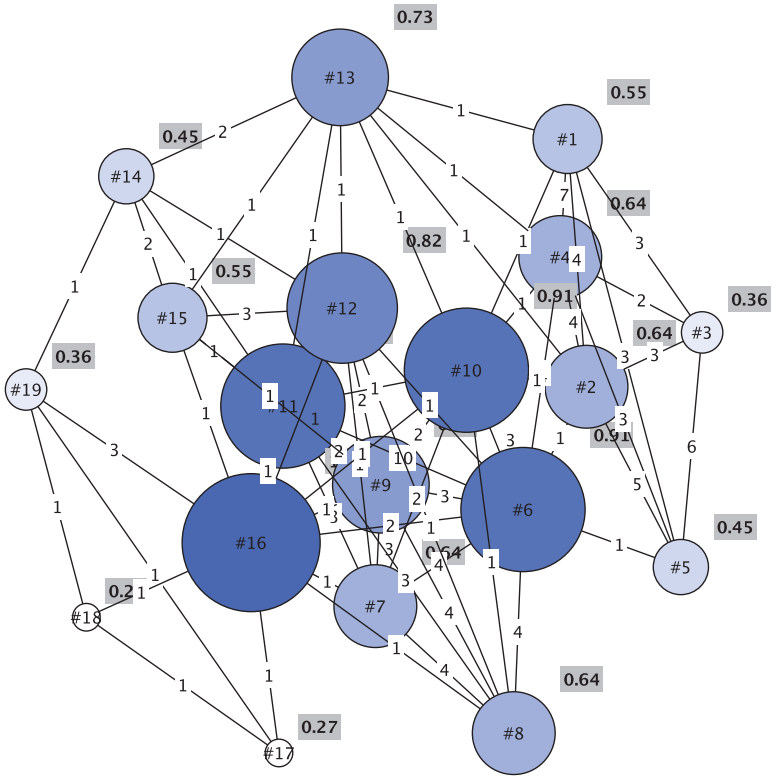

Abbildung 14: gesamtes Netzwerk (Gewichtung: degree-Zentralität)

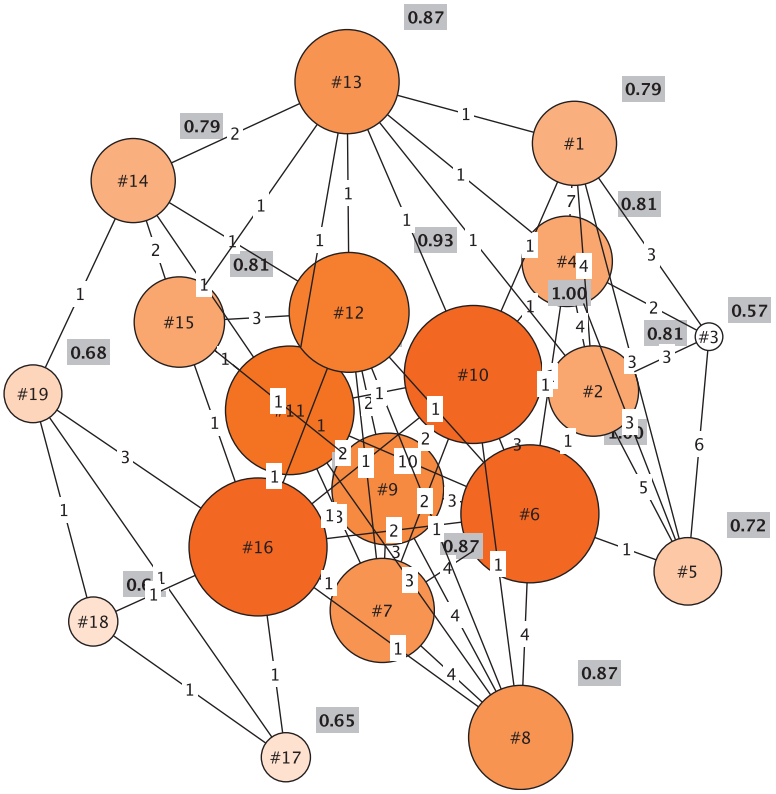

Abbildung 15: gesamtes Netzwerk (Gewichtung: closeness-Zentralität)

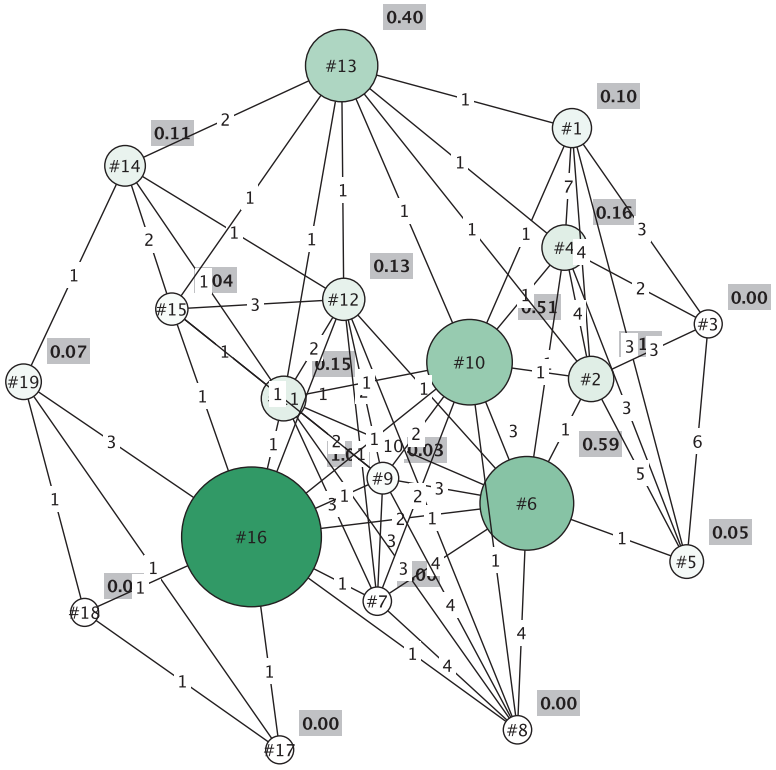

Abbildung 16: gesamtes Netzwerk (Gewichtung: betweeness-Zentralität)

Beschreibung

Die Grafiken 13-16[394] zeigen eine Visualisierung des sozialen Netzwerkes der Attentäter vom 11. September 2001 in New York, Washington D. C. und Pennsylvania. Als Quelle dienten die durch die Washington Post[395] veröffentlichten Ermittlungsergebnisse zu den Kontakten der Attentäter von 2000 bis zum 11. September 2001. Neben einer allgemeinen, ungewichteten Darstellung (Abbildung 13), zeigen die anderen Grafiken eine berechnete und durch die Größe der geometrischen Figuren visualisierte Gewichtung der einzelnen Akteure nach den Zentralitätsmaßen (siehe Kapitel 4.5.1) degree (Abbildung 14), closeness (Abbildung 15) und beweeness (Abbildung 16). In der Tabelle 4 sind die in der SNA verarbeiteten Rohdaten aufgeführt. Hierfür erfolgte zuvor eine Datenextraktion und anwendungsorientierte[396] Transformation der Daten aus der Veröffentlichung der Washington Post.

Liste und Codierung der Akteure (Attentäter 9/11):

#1 Hani Hanjour

#2 Majed Moqed

#3 Nawaf Alhazmi

#4 Salem AlHazmi

#5 Khalid Almihdhar

#6 Mohamed Atta

#7 Waleed M. Alshehri

#8 Wail M. Alshehri

#9 Satam MA Al Suqami

#10 Abdulaziz Alomari

#11 Marwan Al-Shehhi

#12 Fayez Rashid Ahmed Hassan Al Qadi Banihammad

#13 Ahmed Alghamdi

#14 Hamza Alghamdi

#15 Mohand Alshehri

#16 Ziad Samir Jarrah

#17 Ahmed Alnami

#18 Ahmed Ibrahim A. Al Haznawi

#19 Saeed Alghamdi

[394] Abbildung 13-16: eigene Darstellungen.
[395] WEBSTER 2001.
[396] Für die Berechnung und Visualisierung der Zentralitätsmaße wurde die Anwendung yED genutzt.

Tabelle 4: Rohdaten für Beispiel SNA 9/11

Ereignis	A	B	Date
1	#3	#5	2000
2	#3	#5	2000
3	#6	#11	2000
4	#6	#11	2000
5	#14	#15	2001-01
6	#6	#11	2001-02
7	#6	#11	2001-03-11
8	#6	#16	2001-05-02
9	#6	#11	2001-06-13
10	#14	#19	2001-06-15
11	#11	#12	2001-07-01
11	#6	#11	2001-07-01
11	#6	#12	2001-07-01
11	#6	#7	2001-07-01
11	#6	#8	2001-07-01
11	#6	#9	2001-07-01
11	#7	#11	2001-07-01
11	#7	#12	2001-07-01
11	#7	#8	2001-07-01
11	#7	#9	2001-07-01
11	#8	#11	2001-07-01
11	#8	#12	2001-07-01
11	#8	#9	2001-07-01
11	#9	#11	2001-07-01
11	#9	#12	2001-07-01
12	#8	#9	2001-07-03
13	#3	#5	2001-07-04
14	#1	#4	2001-07-29
15	#1	#10	2001-08-02
15	#1	#13	2001-08-02
15	#1	#2	2001-08-02
15	#1	#4	2001-08-02
15	#10	#13	2001-08-02
15	#2	#10	2001-08-02
15	#2	#13	2001-08-02
15	#2	#4	2001-08-02
15	#4	#10	2001-08-02
15	#4	#13	2001-08-02
16	#1	#4	2001-08-06
17	#2	#4	2001-08-25
17	#2	#5	2001-08-25
17	#2	#6	2001-08-25
17	#4	#5	2001-08-25
17	#4	#6	2001-08-25
17	#5	#6	2001-08-25
18	#6	#11	2001-08-26
18	#6	#7	2001-08-26
18	#6	#8	2001-08-26
18	#7	#11	2001-08-26
18	#7	#8	2001-08-26
18	#8	#11	2001-08-26
19	#1	#4	2001-08-26
20	#12	#15	2001-08-27
20	#12	#16	2001-08-27
20	#15	#16	2001-08-27
21	#10	#11	2001-08-28
21	#10	#16	2001-08-28
21	#11	#16	2001-08-28
21	#6	#10	2001-08-28
21	#6	#11	2001-08-28
21	#6	#16	2001-08-28
21	#6	#7	2001-08-28
21	#6	#8	2001-08-28
21	#6	#9	2001-08-28
21	#7	#10	2001-08-28
21	#7	#11	2001-08-28
21	#7	#16	2001-08-28
21	#7	#8	2001-08-28
21	#7	#9	2001-08-28
21	#8	#10	2001-08-28

Ereignis	A	B	Date
21	#8	#11	2001-08-28
21	#8	#16	2001-08-28
21	#8	#9	2001-08-28
21	#9	#10	2001-08-28
21	#9	#11	2001-08-28
21	#9	#16	2001-08-28
22	#1	#4	2001-08-30
23	#1	#2	2001-09-01
23	#1	#3	2001-09-01
23	#1	#4	2001-09-01
23	#1	#5	2001-09-01
23	#2	#3	2001-09-01
23	#2	#4	2001-09-01
23	#2	#5	2001-09-01
23	#3	#4	2001-09-01
23	#3	#5	2001-09-01
23	#4	#5	2001-09-01
24	#1	#2	2001-09-02
24	#1	#3	2001-09-02
24	#1	#5	2001-09-02
24	#2	#3	2001-09-02
24	#2	#5	2001-09-02
24	#3	#5	2001-09-02
25	#2	#5	2001-09-05
26	#16	#19	2001-09-05
27	#6	#11	2001-09-07
28	#16	#19	2001-09-07
29	#12	#15	2001-09-09
29	#9	#12	2001-09-09
29	#9	#15	2001-09-09
30	#6	#11	2001-09-09
31	#6	#10	2001-09-10
32	#13	#14	2001-09-10
33	#1	#2	2001-09-11
33	#1	#3	2001-09-11
33	#1	#4	2001-09-11
33	#1	#5	2001-09-11
33	#2	#3	2001-09-11
33	#2	#4	2001-09-11
33	#2	#5	2001-09-11
33	#3	#4	2001-09-11
33	#3	#5	2001-09-11
33	#4	#5	2001-09-11
34	#6	#10	2001-09-11
34	#6	#7	2001-09-11
34	#6	#8	2001-09-11
34	#6	#9	2001-09-11
34	#7	#10	2001-09-11
34	#7	#8	2001-09-11
34	#7	#9	2001-09-11
34	#8	#9	2001-09-11
34	#9	#10	2001-09-11
35	#11	#12	2001-09-11
35	#11	#13	2001-09-11
35	#11	#14	2001-09-11
35	#11	#15	2001-09-11
35	#12	#13	2001-09-11
35	#12	#14	2001-09-11
35	#12	#15	2001-09-11
35	#13	#14	2001-09-11
35	#13	#15	2001-09-11
35	#14	#15	2001-09-11
36	#16	#17	2001-09-11
36	#16	#18	2001-09-11
36	#16	#19	2001-09-11
36	#17	#18	2001-09-11
36	#17	#19	2001-09-11
36	#18	#19	2001-09-11

Literatur

ABTS, Dietmar/MÜLDER, Wilhelm (2017): *Grundkurs Wirtschaftsinformatik: Eine kompakte und praxisorientierte Einführung*. 9. Aufl. Wiesbaden.

ACKERMANN, Rolf (2005a): Zusammengang von Kriminalistischer Hypothesen-/Versionsbildung und Fallanalyse. In: *Kriminalistik* 7, S. 461–464.

ACKERMANN, Rolf (2005b): Zusammengang von Kriminalistischer Hypothesen-/Versionsbildung und Fallanalyse. In: *Kriminalistik* 8-9, S. 542–544.

ACKERMANN, Rolf (2017): Kriminalistische Fallanalyse. In: *Der rote Faden*. Hrsg. von Horst CLAGES/Rolf ACKERMANN. Heidelberg, S. 131–160.

ACKERMANN, Rolf/CLAGES, Horst/ROLL, Holger (2011): *Handbuch der Kriminalistik: Kriminaltaktik für Praxis und Ausbildung*. 4. Aufl. Stuttgart u. a.

ACKERMANN, Rolf/KORISTKA, Christian et. al. (2000): Zum Stellenwert der Kriminalistik – Teil 3 –: Geschichte der Kriminalistik , Kriminaltaktik und Kriminaltechnik. In: *Kriminalistik* 11, S. 731–736.

ACKERMANN, Rolf/RACHOW, Rudolf (1989): *Untersuchungsplanung in Frage und Antwort*. Berlin.

ACKERMANN, Rolf/STRAUSS, Ernst (1986): *Die kriminalistische Untersuchungsplanung: Untersuchungsmethodik*. Berlin.

AHLF, Enrst-Heinrich (2002): Kriminalpolizeiliche Auswertung/Intelligance. In: *DPolBl* 3, S. 2–5.

AKADEMIE DER POLIZEI HAMBURG (2017): *Modulhandbuch: Studiengang Polizei*. URL: http://akademie-der-polizei.hamburg.de/contentblob/8614246/2e21d9b583510bd53eb7b3c7eecbaa8b/data/modulhandbuch-stand-01-04-2017.pdf (besucht am 29. 12. 2017).

ALBRECHT, Hans-Jörg/GRAFE, Adina/KILCHLING, Michael (2008): *Rechtswirklichkeit der Auskunftserteilung über Telekommunikationsverbindungsdaten nach §§ 100g, 100h StPO: Forschungsbericht im Auftrag des Bundesministeriums der Justiz*. Freiburg. URL: https://www.hintergrund.de/wp-content/uploads/2008/01/mpi-gutachten.pdf (besucht am 13. 01. 2018).

APEL, Detlef/BEHME, Wolfgang (2010): Datenintegration – Ein Prozess zur Verbesserung der Datenqualität. In: *Analytische Informationssysteme*. Hrsg. von Peter CHAMONI/Peter GLUCHOWSKI. Berlin und Heidelberg, S. 115–130.

APEL, Detlef/BEHME, Wolfgang et. al. (2015): *Datenqualität erfolgreich steuern: Praxislösungen für Business-Intelligence-Projekte*. 3. Aufl. Heidelberg.

BERTIN, Jacques (1982): *Graphische Darstellungen und die graphische Weiterverarbeitung der Information*. Berlin und New York.

BITCOM (2014): *Big-Data-Technologien – Wissen für Entscheider: Leitfaden*. URL: https://www.bitkom.org/noindex/Publikationen/2014/Leitfaden/

Big-Data-Technologien-Wissen-fuer-Entscheider/140228-Big-Data-Technologien-Wissen-fuer-Entscheider.pdf (besucht am 08. 01. 2018).

BITCOM (o. D.): *Arbeitskreis Big Data und Advanced Analytics*. URL: https://www.bitkom.org/Bitkom/Organisation/Gremien/Big-Data-und-Advanced-Analytics.html (besucht am 08. 01. 2018).

BLEIHOLDER, Jens/SCHMID, Joachim (2015): Datenintegration und Deduplizierung. In: *Daten- und Informationsqualität*. Hrsg. von Knut HILDEBRAND et. al. Wiesbaden, S. 121–140.

BRISACH, Carl-Ernst (1992): Kriminalistische Handlungslehre. In: *Kriminalistik*. Hrsg. von Edwin KUBE/Hans Udo STÖRZER/Klaus Jürgen TIMM. Stuttgart u. a., S. 167–197.

BUDDENBROCK, Andrea von (2015): Der Einsatz von Mantrailern bei Kapitaldelikten. In: *Kriminalitätsbekämpfung – ein Blick in die Zukunft*. Hrsg. von Heiko ARTKÄMPER/Horst CLAGES. Stuttgart u. a., S. 41–59.

BUNDESKRIMINALAMT (2015): *Täter im Bereich Cybercrime: Eine Literaturanalyse*. Wiesbaden. URL: https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/Forschungsergebnisse/2015TaeterImBereichCybercrime.html (besucht am 13. 01. 2018).

BUNDESNETZAGENTUR FÜR ELEKTRIZITÄT, GAS, TELEKOMMUNIKATION, POST UND EISENBAHNEN, Hrsg. (2017): *Technische Richtlinie zur Umsetzung gesetzlicher Maßnahmen zur Überwachung der Telekommunikation, Erteilung von Auskünften (TR TKÜV)*. URL: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Anbieterpflichten/OeffentlicheSicherheit/TechnUmsetzung110/Downloads/TR%20TKUEV%20Version%207.0%20pdf%20deutsch.pdf?__blob=publicationFile&v=1 (besucht am 13. 01. 2018).

BURGHARD, Waldemar (1993): *Täterermittlung/Fahndung*. 3. Aufl. Hilden.

BURKHARD, Remo Aslak (2008): Informationsarchitektur. In: *Kompendium Informationsdesign*. Hrsg. von Wibke WEBER. Berlin, Heidelberg, S. 303–319.

CIVELLI, Ignaz (2010a): Analysepsychologie: Teil 1: Psychologische Faktoren der Wahrnehmungsbeeinflussung beim Polizeianalysten. In: *Kriminalistik* 11, S. 665–671.

CIVELLI, Ignaz (2010b): Aussagepsychologie: Teil 2: Kognitive Verzerrungen – Wahrnehmungsfallen und Urteilsfehler bei Polizeianalysten. In: *Kriminalistik* 12, S. 719–725.

CIVELLI, Ignaz (2010c): Der Analysezyklus als Tool für Lageanalysen. In: *POLIZEIheute* 4, S. 131–136.

CLAGES, Horst (2017a): Aktenführung im strafprozessualen Ermittlungsverfahren. In: *Der rote Faden*. Hrsg. von Horst CLAGES/Rolf ACKERMANN. Heidelberg, S. 63–82.

CLAGES, Horst (2017b): Einführung in die Kriminalistik. In: *Der rote Faden*. Hrsg. von Horst CLAGES/Rolf ACKERMANN. Heidelberg, S. 1–22.

CLAGES, Horst (2017c): Polizeiliche Fahndung. In: *Der rote Faden*. Hrsg. von Horst CLAGES/Rolf ACKERMANN. Heidelberg, S. 265–287.

CLAGES, Horst/ACKERMANN, Rolf, Hrsg. (2017): *Der rote Faden: Grundsätze der Kriminalpraxis*. 13. Aufl. Heidelberg.

DÄUBLER, Wolfgang et. al. (2014): *Bundesdatenschutzgesetz: Kompaktkommentar zum BDSG*. 4. Aufl. Frankfurt/Main.

DORSCHEL, Joachim (2015): Einführung und Überblick. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 5–13.

DÜRR, Holger (2004): Anwendungen des Data Mining in der Praxis. Diss. Ulm: Universität Ulm. URL: http://www.mathematik.uni-ulm.de/sai/ws03/dm/arbeit/duerr.pdf (besucht am 08. 01. 2018).

EAGLE, Nathan/PENTLAND, Alex/LAZER, David (2009): Inferring Social Network Structure using Mobile Phone Data. In: *Proc. of National Academy of Sciences*, S. 15274–15278. URL: https://www.researchgate.net/publication/253153146 (besucht am 13. 01. 2018).

ECKERSON, Wayne W. (2002): *DATA QUALITY AND THE BOTTOM LINE: Achieving Business Success through a Commitment to High Quality Data*. URL: http://download.101com.com/pub/tdwi/Files/DQReport.pdf (besucht am 11. 12. 2017).

EDER, Gerald (2004): Professionelles IT-"Werkzeug" für die Ermittler der bayrischen Landespolizei: Computersysteme können die Arbeit der Beamten erheblich erleichtern. In: *Homeland Security* 3, S. 4–6.

EDER, Gerald (2005): EASy – ein professionelles und etabliertes Werkzeug zur Fallbearbeitung, Auswertung und Analyse. In: *der kriminalist* 10, S. 387–391.

ELIAS, Norbert (2009): *Was ist Soziologie?* 11. Aufl. Weinheim und München.

ESCH, Heike (2009): Massendatenanalyse in der Praxis: Die Arbeitsweise der Verfahrensbegleitenden Analyse im Hessischen LKA am Beispiel eines Tötungsdeliktes. In: *Kriminalistik* 12, S. 680–684.

ESRI (2008): *Law Enforcement: GIS Solutions for Proactive Policing and Informed Response*. URL: http://www.esri.com/library/brochures/pdfs/lawenforcement.pdf (besucht am 08. 01. 2018).

FARKISCH, Kiumars (2011): *Data-Warehouse-Systeme kompakt: Aufbau, Architektur, Grundfunktionen*. Berlin, Heidelberg.

FASEL, Daniel/MEIER, Andreas (2016): Was versteht man unter Big Data und NoSQL? In: *Big Data*. Hrsg. von Daniel FASEL/Andreas MEIER. Wiesbaden, S. 3–16.

FELDEN, Carsten (2006): Text Mining als Anwendungsbereich von Business Intelligence. In: *Analytische Informationssysteme*. Hrsg. von Peter CHAMONI/Peter GLUCHOWSKI. Berlin und Heidelberg, S. 283–304.

FÖHL, Ulrich/THEOBALD, Elke (2015): Big Data und Electronic Commerce – Neue Erkenntnisse zur Customer Journey. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 123–133.

FRIEDL, Jeffrey E. F. (2008): *Reguläre Ausdrücke*. 3. Aufl. Beijing u. a.

GADATSCH, Andreas/LANDROCK, Holm (2017): *Big Data für Entscheider: Entwicklung und Umsetzung datengetriebener Geschäftsmodelle*. Wiesbaden.

GARFINKEL, Simson L. (2009): Big Brother mit Sehschwäche. In: *Spektrum der Wissenschaft* 10, S. 90–95.

GAUS, Wilhelm (2003): *Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval*. 4. Aufl. Berlin und Heidelberg.

GEISELBERGER, Heinrich/MOORSTEDT, Tobias, Hrsg. (2013): *Big Data: Das neue Versprechen der Allwissenheit*. Berlin.

GOLA, Peter et. al. (2015): *Bundesdatenschutzgesetz*. 12. Aufl. München.

GRUTZPALK, Jonas (2013): Zur Erforschung des Wissensmanagements in Sicherheitsbehörden. In: *der kriminalist* 6, S. 21–23.

HABERBERGER, Evi/TALARCZYK, Harald (2007): Elektronische Spurensuche in Mobilfunkdaten. In: *der kriminalist* 9, S. 367–371.

HAKE, Günter/GRÜNREICH, Dietmar/MENG, Liqiu (2002): *Kartographie: Visualisierung raum-zeitlicher Informationen*. 8. Aufl. Berlin.

HANSEN, Hans Robert/MENDLING, Jan/NEUMANN, Gustaf (2015): *Wirtschaftsinformatik: Grundlagen und Anwendungen*. 11. Aufl. Berlin u.a.

HÄUSSLING, Roger (2010): Relationale Soziologie. In: *Handbuch Netzwerkforschung*. Hrsg. von Christian STEGBAUER/Roger HÄUSSLING. Wiesbaden, S. 63–87.

HEINRICH, Lutz J./RIEDL, René/STELZER, Dirk (2014): *Informationsmanagement: Grundlagen, Aufgaben, Methoden*. 11. Aufl. Berlin und Boston.

HEINRICHS, Axel/WILHELM, Jörg (2010): *Funkzellenauswertung: Rechtliche und taktische Aspekte der telekommunikativen Spurensuche*. URL: http://www.krimmalpolizei.de/articles,funkzenenauswertung,1,275,prmt.htm.

HELD, Heiko (2003): *DataMining in der polizeilichen Anwendung*. URL: https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/ForumKI/ForumKI12003/kiforum2003HeldLangfassung.html (besucht am 08. 01. 2018).

HENNERMANN, Karl/WOLTERING, Manuel (2014): *Kartographie und GIS: Eine Einführung*. 2. Aufl. Darmstadt.

HENTSCHEL, Manfred W./PÖTZL, Norbert F. (1986): „Die Position der RAF hat sich verbessert": Der ehemalige BKA-Chef Horst Herold über Terroristen und

Computer-Fahndung. In: *DER SPIEGEL* 37, S. 38–61. URL: http://www.spiegel.de/spiegel/print/d-13519259.html (besucht am 13. 01. 2018).

HILDEBRAND, Knut et. al. Hrsg. (2015): *Daten- und Informationsqualität: Auf dem Weg zur Information Excellence*. 3. Aufl. Wiesbaden.

HIPP, Jochen (2003): *Wissensentdeckung in Datenbanken mit Assoziationsregeln: Verfahren zur effizienten Regelgenerierung und deren Integration in den Wissensentdeckungsprozess: Zugl. Diss. der Fakultät für Informations- und Kognitionswissenschaften der Eberhard-Karls-Universität Tübingen zur Erlangung des Grades eines Doktors der Naturwissenschaften*. Tübingen.

HORVATH, Sabine (2013): *Aktueller Begriff: Big Data*. URL: https://www.bundestag.de/blob/194790/c44371b1c740987a7f6fa74c06f518c8/big_data-data.pdf (besucht am 08. 01. 2018).

HUFNAGEL, Sven/KOLLMANN, Wulf (2015): Analysen von großen und komplexen Datenmengen. In: *Kriminalitätsbekämpfung – ein Blick in die Zukunft*. Hrsg. von Heiko ARTKÄMPER/Horst CLAGES. Stuttgart u. a., S. 139–150.

HUMMELTENBERG, Wilhelm (2010): Vom Content Management zum Enterprise Decision Management – eine Chronologie der Business Intelligence-Systeme. In: *Analytische Informationssysteme*. Hrsg. von Peter CHAMONI/Peter GLUCHOWSKI. Berlin und Heidelberg, S. 17–36.

IBM (2012): *IBM i2 Analyst's Notebook Social Network Analysis*. Somers, NY. URL: https://cryptome.org/2013/12/ibm-i2-sna.pdf (besucht am 13. 01. 2018).

INMON, William H. (2005): *Building the Data Warehouse*. 4. Aufl. Indianapolis.

JAEGER, Rolf Rainer (2005): Die Fesselung des Kriminalisten an seinem Computer-Arbeitsplatz: Chancen der Entfesselung durch Unterstützung qualifizierter Angestellter. In: *der kriminalist* 10, S. 415–422.

JANSEN, Dorothea/DIAZ-BONE, Rainer (2011): Netzwerkstrukturen als soziales Kapital: Konzepte und Methoden zur Analyse struktureller Einbettung. In: *Soziale Netzwerke*. Hrsg. von Johannes WEYER. München, S. 73–108.

JÜNGLING, Thomas (2011): *Dein Handy weiß, wo du letzten Monat warst*. URL: https://www.welt.de/print/die_welt/wirtschaft/article13247752/Dein-Handy-weiss-wo-du-letzten-Monat-warst.html (besucht am 13. 01. 2018).

KÄNEL, Siegfried von (o. D.): *DAA Wirtschafts-Lexikon: Datenprozesse*. URL: https://media.daa-pm.de/ufv_wirtschaftslexikon/Html/D/PDF/Datenprozesse.pdf (besucht am 13. 01. 2018).

KAY, Wolfgang (2013): Datenschutz – Zweckänderung der Datennutzung. In: *Kriminalistik* 7, S. 495–503.

KEUPER, Frank/SCHMIDT, Dietmar/SCHOMANN, Marc, Hrsg. (2014): *Smart Big Data Management*. Berlin.

KING, Stefanie (2014): *Big Data: Potential und Barrieren der Nutzung im Unternehmenskontext: Zugl. Diss. Universität Innsbruck, 2013*. Wiesbaden.

KLIMETZEK, Christian (2013): *Konzeption von Data Mining-Prozessen im Kontext von massiv parallelen Architekturen: Zugl. Inauguraldiss. zur Erlangung der Würde eines Doktors der Wirtschaftswissenschaft der Fakultät für Wirtschaftswissenschaft der Ruhr-Universität Bochum*. Aachen.

KOHLHAMMER, Jörn/FROFF, Dirk U./WIENER, Andreas (2016): Der Markt für Visual Business Analytics. In: *Analytische Informationssysteme*. Hrsg. von Peter GLUCHOWSKI/Peter CHAMONI. Berlin und Heidelberg, S. 303–323.

KÖRFFER, Barbara (2014): Auswertung personenbezogener Daten für Strafverfolgung und Gefahrenabwehr – genügen die gesetzlichen Grundlagen zum Schutz des Rechts auf informationeile Selbstbestimmung? In: *DANA – Datenschutz Nachrichten* 4, S. 146–150.

KRCMAR, Helmut (2015): *Einführung in das Informationsmanagement*. 2. Aufl. Berlin und Heidelberg.

KREMPEL, Lothar (2005): *Visualisierung komplexer Strukturen: Grundlagen der Darstellung mehrdimensionaler Netzwerke*. Frankfurt/Main und New York.

KREMPL, Stefan (2017): *Bundestagsgutachten: Neue Vorratsdatenspeicherung ist rechtswidrig*. URL: https://www.heise.de/newsticker/meldung/Bundestagsgutachten-Neue-Vorratsdatenspeicherung-ist-rechtswidrig-3617806.html (besucht am 08. 01. 2018).

KRIEGER, Winfried et. al. (o. D.): Kennzahlen. In: *Gabler Wirtschaftslexikon*. Hrsg. von SPRINGER GABLER VERLAG. URL: http://wirtschaftslexikon.gabler.de/Archiv/54801/kennzahlen-v11.html (besucht am 13. 01. 2018).

KUBE, Edwin/SCHREIBER, Manfred (1992): Theoretische Kriminalistik. In: *Kriminalistik*. Hrsg. von Edwin KUBE/Hans Udo STÖRZER/Klaus Jürgen TIMM. Stuttgart u. a., S. 1–17.

LACKES, Richard/SIEPERMANN, Markus (o. D.[a]): HTML. In: *Gabler Wirtschaftslexikon*. Hrsg. von SPRINGER GABLER VERLAG. URL: http://wirtschaftslexikon.gabler.de/Archiv/75639/html-v10.html (besucht am 06. 12. 2017).

LACKES, Richard/SIEPERMANN, Markus (o. D.[b]): Web 2.0. In: *Gabler Wirtschaftslexikon*. Hrsg. von SPRINGER GABLER VERLAG. URL: http://wirtschaftslexikon.gabler.de/Archiv/80667/web-2-0-v10.html.

LACKES, Richard/SIEPERMANN, Markus (o. D.[c]): XML. In: *Gabler Wirtschaftslexikon*. Hrsg. von SPRINGER GABLER VERLAG. URL: http://wirtschaftslexikon.gabler.de/Archiv/75488/xml-v12.html (besucht am 06. 12. 2017).

LANQUILLON, Carsten/MALLOW, Hauke (2015a): Advanced Analytics mit Big Data. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 55–89.

LANQUILLON, Carsten/MALLOW, Hauke (2015b): Big Data-Lösungen. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 263–277.

LANQUILLON, Carsten/MALLOW, Hauke (2015c): Grenzen konventioneller Business-Intelligence-Lösungen. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 255–263.

LANZINGER, Robert/KELLNER, Christian (2002): Bekämpfung der Schleusungskriminalität unter dem Aspekt des "Neuen Auswerteverständnisses". In: *DPolBl* 5, S. 17–20.

LISBACH, Bertrand (2011): *Linguistisches Identity Matching: Paradigmenwechsel in der Suche und im Abgleich von Personendaten*. Wiesbaden.

LUDWIG, Joachim (2008): Auswertung von Telekommunikationsdaten. In: *der kriminalist* 6, S. 252–258.

MAINZER, Klaus (2014): *Die Berechnung der Welt: Von der Weltformel zu Big Data*. München.

MANGOLD, Roland (2008): Informationspsychologie. In: *Kompendium Informationsdesign*. Hrsg. von Wibke WEBER. Berlin, Heidelberg, S. 253–271.

MERTENS, Peter/BODENDORF, Freimut et. al. (2017): *Grundzüge der Wirtschaftsinformatik*. 12. Aufl. Berlin.

MERTENS, Peter/WIECZORREK, Hans Wilhelm (2000): *Data X Strategien: Data Warehouse, Data Mining und operationale Systeme für die Praxis*. Berlin, Heidelberg und New York.

MEYER-GOSSNER, Lutz/SCHMITT, Bertram (2015): *Strafprozessordnung: Gerichtsverfassungsgesetz, Nebengesetze und ergänzende Bestimmungen*. 58. Aufl. München.

MORBEY, Guilherme (2011): *Datenqualität für Entscheider in Unternehmen: Ein Dialog zwischen einem Unternehmenslenker und einem DQ-Experten*. Wiesbaden.

MÜLLER, Stefan (2016): Erweiterung des Data Warehouse um Hadoop, NoSQL & Co. In: *Big Data*. Hrsg. von Daniel FASEL/Andreas MEIER. Wiesbaden, S. 139–158.

MÜNCH, Holger (2017): *Polizei im Umbruch – Herausforderungen und Zukunftsstrategien: Polizeiliche Herausforderungen und Zukunftsstrategien aus Sicht des Bundeskriminalamtes*. URL: https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Herbsttagungen/2017/herbsttagung2017MuenchLangfassung.pdf?__blob=publicationFile&v=4 (besucht am 15. 01. 2018).

MUSOLFF, Cornelia/HOFFMANN, Jens (2006): *Täterprofile bei Gewaltverbrechen: Mythos, Theorie, Praxis und forensische Anwendung des Profilings*. 2. Aufl. Heidelberg.

NACK, Armin (1999): Beweisrecht: Zum Beweiswert kriminaltechnischer Untersuchungsergebnisse. In: *Kriminalistik* 1, S. 32–39.

NEUHAUS, Ralf/ARTKÄMPER, Heiko (2014): *Kriminaltechnik und Beweisführung im Strafverfahren*. München.

NEWMANN, Daniel (2017): *Realizing The Potential Of Big Data And Analystics*. URL: https://www.forbes.com/sites/danielnewman/2017/01/31/realizing-the-potential-of-big-data-and-analytics/2/#289aa26368b1 (besucht am 08. 01. 2018).

NITSCHKE, Christian/ROLLINGER, Christian (2015): „Network Analysis is performed.“: Die Analyse sozialer Netzwerke in den Altertumswissenschaften: Rückschau und aktuelle Forschungen. In: *Knoten und Kanten III*. Hrsg. von Markus GAMPER/Linda RESCHKE/Marten DÜRING. Bielefeld, S. 213–259.

NONNINGER, Dirk (2002): Analyse – eine "neue" Form der polizeilichen Informationsverarbeitung. In: *DPolBl* 5, S. 6–8.

OMRI, Fouad (2015): Big Data-Analysen: Anwendungsszenarien und Trends. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 104–112.

OTTO, Boris/ÖSTERLE, Hubert (2016): *Corporate Data Quality: Voraussetzung erfolgreicher Geschäftsmodelle*. Berlin und Heidelberg.

PIAZZA, Franca (2010): *Data Mining im Personalmanagement: Eine Analyse des Einsatzpotenzials zur Entscheidungsunterstützung: Zugl. Diss. Universität des Saarlandes, 2009*. Wiesbaden.

PIRO, Andrea/GEBAUER, Marcus (2015): Definition von Datenarten zur konsistenten Kommunikation im Unternehmen. In: *Daten- und Informationsqualität*. Hrsg. von Knut HILDEBRAND et. al. Wiesbaden, S. 141–154.

PROBST, Gilbert/RAUB, Steffen/ROMHARDT, Kai (2012): *Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen*. 7. Aufl. Wiesbaden.

RAHM, Erhard/SAAKE, Gunter/SATTLER, Kai-Uwe (2015): *Verteiltes und Paralleles Datenmanagement: Von verteilten Datenbanken zu Big Data und Cloud*. Berlin und Heidelberg.

REEZ, Norbert (2007): Perspektiven der polizeilichen Auswertung und Analyse. In: *DPolBl* 1, S. 13–19.

REICHERTZ, Jo/WILZ, Sylvia Marlene (2016): Polizeiliche Aufklärungsarbeit 2.0. In: *SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis* 1, S. 31–39.

REITERER, Harald/JETTER, Hans-Christian (2013): Informationsvisualisierung. In: *Grundlagen der praktischen Information und Dokumentation*. Hrsg. von Rainer KUHLEN/Wolfgang SEMAR/Dietmar STRAUCH. Berlin und Boston, S. 192–206.

REUSS, Andreas/ZWIESLER, Hans-Joachim (2005): *Ein generisches Kreislaufmodell zur Einbettung von Data-Mining-Analysen in die Geschäftsprozesse von Unternehmen – mit einem Fallbeispiel aus der Versicherung*. URL: https://www.researchgate.net/publication/267563485_Ein_generisches_Kreislaufmodell_zur_Einbettung_von_Data-Mining-Analysen_in_die_Geschaftsprozesse_von_Unternehmen_-_mit_einem_Fallbeispiel_aus_der_Versicherung (besucht am 08. 01. 2018).

ROERICH, Jonathan (2017): Die Neuregelung der Funkzellenabfrage. In: *Kriminalistik* 3, S. 175–181.

ROLFES, Manfred (2017): Predictive Policing: Beobachtungen und Reflexionen zur Einführung und Etablierung einer vorhersagenden Polizeiarbeit. In: *Geoinformation & Visualisierung: Festschrift anlässlich der Emeritierung von Herrn Prof. Dr. Hartmut Asche im März 2017*. Hrsg. von FACHGRUPPE GEOINFORMATIK DES INSTITUTS FÜR GEOGRAPHIE DER UNIVERSITÄT POTSDAM. Potsdam, S. 51–76.

ROLL, Holger (2013): Kriminalistische Informationsbewertung. In: *Kriminalistik gestern-heute-morgen*. Hrsg. von Heiko ARTKÄMPER/Horst CLAGES. Stuttgart u. a., S. 361–396.

ROLL, Holger (2017): Kriminalistische Tatortarbeit. In: *Der rote Faden*. Hrsg. von Horst CLAGES/Rolf ACKERMANN. Heidelberg, S. 82–130.

SCHAAR, Peter (2002): *Datenschutz im Internet: Die Grundlagen*. München.

SCHNIDER, Dani et. al. (2016): *Data Warehouse Blueprints: Business Intelligence in der Praxis*. München.

SCHULMEYER/CHRISTIAN (2015): Big Data-Analysen auf Basis technischer Methoden und Systeme. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 307–330.

SCHULZKI-HADDOUTI, Christiane (2011): Gläserne soziale Netzwerke: Fahndung in digitalen sozialen Interaktionen. In: *Bürgerrechte & Polizei/CILIP* 1, S. 32–39.

SCHUMANN, Heidrun/MÜLLER, Wolfgang (2000): *Visualisierung: Grundlagen und allgemeine methoden*. Berlin, Heidelberg und New York.

SCHWARZ, Clemenz (2015): Digitale Daten im persönlichen Umfeld – Faktenbasierte Entscheidungsfindung im Bereich der Sicherheit. In: *Kriminalitätsbekämpfung – ein Blick in die Zukunft*. Hrsg. von Heiko ARTKÄMPER/Horst CLAGES. Stuttgart u. a., S. 287–315.

SEIDEL, Ludwig Michael (2013): *Text Mining als Methode zur Wissensexploration: Konzepte, Vorgehensmodelle, Anwendungsmöglichkeiten*. Wismar. URL: http://www.wi.hs-wismar.de/~cleve/vorl/projects/da/13-Master-Seidel.pdf (besucht am 08. 01. 2018).

SOUKUP, Otmar/BARTEN, Wolfgang (2013): Die Taten des "Nationalsozialistischen Untergrundes (NSU)": Herausforderungen an die BAO des Bundeskriminalamtes. In: *Kriminalistik* 1, S. 22–24.

SPANG, Thomas (2008): Kriminaltaktik, kriminalistisches Denken und kriminalistische Methode. In: *Grundlagen der Kriminalistik / Kriminologie*. Hrsg. von Horst CLAGES/Klaus NEIDHARDT/Robert WEIHMANN. Hilden, S. 61–80.

SPRANGER, Michael/LABUDDE, Dirk (2017): Textforensik. In: *Forensik in der digitalen Welt*. Hrsg. von Dirk LABUDDE/Michael SPRANGER. Berlin und Heidelberg, S. 167–198.

STEGBAUER, Christian/RAUSCH, Alexander (2013): *Einführung in NetDraw: Erste Schritte mit dem Netzwerkvisualisierungsprogramm*. Wiesbaden.

STOCK, Oliver (2017): KNIME – bitte was oder wie? In: *proPolizei* 3, S. 20–21. URL: www.polizei-nds.de/download/72929/proPOLIZEI_Ausgabe_Mai_Juni_2017.pdf (besucht am 13. 01. 2018).

SWAN, Siegfried (2003): *Psychologische Grundlagen der nachrichtendienstlichen Auswertung*. Brühl. URL: http://psydok.psycharchives.de/jspui/bitstream/20.500.11780/137/1/beitraege_is_23.pdf (besucht am 21. 09. 2017).

THEOBALD, Elke/FÖHL, Ulrich (2015): Big Data wird zu Smart Data – Big Data in der Marktforschung. In: *Praxishandbuch Big Data*. Hrsg. von Joachim DORSCHEL. Wiesbaden, S. 112–123.

TITTMANN, Peter (2011): *Graphentheorie: Eine anwendungsorientierte Einführung*. 2. Aufl. München.

VÖLKER, Rainer/SAUER, Sigrid/SIMON, Monika (2007): *Wissensmanagement im Innovationsprozess*. Heidelberg.

VRIES, Hinrich de (2010): Ist die Kriminalistik eine Wissenschaft? In: *SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis* 3, S. 27–35. URL: http://dx.doi.org/10.7396/2010_3_C. (besucht am 19. 01. 2018).

WAGNER, Ralf/MICHAELI, Rainer (2009): Competitive Intelligence. In: *Sicherheit in Organisationen*. Hrsg. von Sven Max LITZCKE/Helmut MÜLLER-ENBERGS. Frankfurt/Main, S. 13–31.

WALDER, Hans/HANSJAKOB, Thomas (2016): *Kriminalistisches Denken*. 10. Aufl. Heidelberg.

WEBSTER, Bill (2001): *The Plot: A Web of Connections*. Hrsg. von THE WASHINGTON POST. URL: www.washingtonpost.com/wp-srv/nation/graphics/attack/investigation_24.html (besucht am 31. 01. 2018).

WEICHERT, Thilo (o.D.): *Big Data und Datenschutz*. URL: https://www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf.

WEICHERT, Thilo (2013): Big Data – eine Herausforderung für den Datenschutz. In: *Big Data*. Hrsg. von Heinrich GEISELBERGER/Tobias MOORSTEDT. Berlin, S. 131–148.

WEIHMANN, Robert/VRIES, Hinrich de (2014): *Kriminalistik*. 13. Aufl. Hilden.

WEINER, Andreas M. (2005): Data Mining auf Datenströmen. Diss. URL: `http://lgis.informatik.uni-kl.de/archiv/wwwdvs.informatik.uni-kl.de/courses/seminar/SS2005/Ausarbeitung15.pdf` (besucht am 08. 01. 2018).

WESTPHAL, Norbert (2010): *Der kriminalistische Beweis*. Hilden.

WEYER, Johannes (2011a): Netzwerke in der mobilen Echtzeit-Gesellschaft. In: *Soziale Netzwerke*. Hrsg. von Johannes WEYER. München, S. 3–38.

WEYER, Johannes (2011b): Zum Stand der Netzwerkforschung in den Sozialwissenschaften. In: *Soziale Netzwerke*. Hrsg. von Johannes WEYER. München, S. 39–69.

WEYER, Johannes/FINK, Robin D./LIBSCHIK, Tobias (2011): Softwarebasierte Methoden der Netzwerk-Analyse. In: *Soziale Netzwerke*. Hrsg. von Johannes WEYER. München, S. 109–131.

WISSENSCHAFTLICHE KOMMISSION WIRTSCHAFTSINFORMATIK UND GESELLSCHAFT FÜR INFORMATIK E.V. (2007): Mitteilungen der Wissenschaftlichen Kommission Wirtschaftsinformatik und des GI-Fachbereichs Wirtschaftsinformatik. In: *WIRTSCHAFTSINFORMATIK* 49.4, S. 318–325.

WOJCIECHOWSKI, Krystian (2012): Einführung in die Analyse von Daten aus sozialen Netzwerken. In: *mepa*, S. 49–52.

ZIERCKE, Jörg (2014): Kriminalistik 2.0: Effektive Strafverfolgung im Zeitalter des Internet aus Sicht des BKA. In: *Kriminalistik* 1, S. 10–17.

ZWIRNER, Marcus (2015): Datenbereinigung zielgerichtet eingesetzt zur permanenten Datenqualitätssteigerung. In: *Daten- und Informationsqualität*. Hrsg. von Knut HILDEBRAND et. al. Wiesbaden, S. 101–120.