Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von Dipl. Inform. Stefan Kiltz

geb. am 14.07.1973 in Magdeburg

Gutachterinnen/Gutachter

Prof. Dr.-Ing. Jana Dittmann
Prof. Eoghan Casey
Prof. Sabah Jassim

Magdeburg, den 16.07.2020

Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics

Abstract

With the widespread use of IT systems, those systems became both targets for attacks and were used as tools to stage attacks and therefore are subject to forensic examinations. Forensic sciences themselves allow questions about the accuracy of inferences made by forensic scientists. Loss, error, uncertainties about measurements and inferences need to be indicated and studies must be made to enable the estimation of these as demanded in literature. Based on those facts, in this thesis we set ourselves the research question of whether a data-centric approach can be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce error, loss and uncertainty.

We want to apply our research to both digital and digitised forensics with the latter being concerned with the usage of IT systems to support crime scene forensics. Deliberately setting aside the association process and the event reconstruction, we are solely concerned with digital data contained in IT systems and its usage during forensic examinations.

We also use the notion of forensic examination in the broadest possible term, ranging from technical support/troubleshooting all the way up to court cases. However, we maintain the stringent demands of comprehensibility of the whole examination and the application of scientific methods. We also use the notion of IT system in the broadest possible term and do not restrict ourselves to desktop and server IT but only expect our centre of interest to process digital data, thus ranging from embedded systems to cloud computing. From the viewpoint of digital forensics, this heterogeneity and vastness poses the challenge of finding a common description of processed data and processing functionality for use in forensics.

We devise and contribute to the scientific community a five-step methodology resulting in our Data-Centric Examination Approach DCEA. We firstly formally describes loss, error and uncertainty regarding data contained in IT systems based on a modelling of the relationship of all data ever available, data used in all forensic investigations ever and the case-specific data for a given incident. Secondly we apply the selected characteristics of the ISO/IEC 7498 model (commonly known as ISO/OSI reference model) to data stored, processed, communicated in IT systems and construct layers of data by giving them semantics that support the forensic process to distinguish forensic data types for digital (6 in total) and digitised forensics (10 in total). Thirdly we use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process for digital forensics (6 in total) and digitised forensics (10 in total), which include tools and toolkits, based on an existing model of transfer functions. Fourthly, we use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics (6 in total). We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics (6 in total). Finally, we use our forensic data types, sets of methods and sets of examination steps for the forensic process to provide a qualitative estimation on loss, error and uncertainty in forensic examinations based on the presence, absence or diversity of forensic data types.

We test our Data-Centric Examination Approach DCEA on three non-standard, actively researched topics in digital and digitised forensics, where the examination description is even more challenging. We use existing scientific research reports for systems used in video surveillance in

digital forensics and digitised forensic dactyloscopy in digitised forensics. We also conduct previously unpublished research on processor-controlled components for digital forensics. In our experiments for all use cases in total we evaluated 31 methods from the sets of methods of the forensic process for both digital and digitised forensic. We covered all 6 examination steps and detected estimations for loss on 3 occasions, for error on 1 occasion and uncertainty on 2 occasions.

Our main findings also return important requirements for the application of the Data-Centric Examination Approach DCEA, namely the conduction of a system landscape analysis for an estimation regarding the forensic data types likely to be contained in the system under examination and thus being recoverable (at least in theory) and to determine system boundaries and the systems used for the subsequent investigation, analysis and documentation. Further, we require the context-sensitive definition of sets of examination steps, sets of methods for the forensic process and forensic data types according to the application area. Crucial, as shown in the previously unpublished research is the level of detail selection together with its justification, which that set the boundaries for the accuracy for the qualitative estimates towards loss, error and uncertainty for a given forensic examination.

In conducting our research, we arrive with further contributions. Using the sets of examination steps, sets of methods for the forensic process and the forensic data types, the Data-Centric Examination Approach DCEA provides a common language to describe the data and their processing using methods of the forensic process and result data and their composition ordered in time and space by the examination steps.

Our approach, given matching forensic data type definitions, allows for the intra-examination comparison for the methods used in the examination that could be used as one decision criterion for the selection of a specific method or as a support for tool testing.

If additionally also the sets of methods for the forensic process and sets of examination steps match, even an inter-examination comparison is possible that can be supportive in a comparative evaluation of the degree of maturity of the examinations or in questions regarding the evidentiary value.

This final version of the thesis addresses the very helpful and stimulating questions and comments raised in the reviews and the colloquium by my thesis advisor Jana Dittmann and my reviewers Eoghan Casey and Sabah Jassim and for which we are very thankful, indeed.

Zusammenfassung

Durch die alle Lebensbereiche durchdringende Anwendung von IT -Systemen wurden diese auch das Ziel von Angriffen und wurden für die Angriffsdurchführung verwendet. Dies bedingt forensische Untersuchungen zur Vorfallsaufklärung. Die forensischen Wissenschaften befürworten Fragen zur Genauigkeit von Deduktionen, welche von forensischen Experten getätigt werden. Verluste, Fehler und Unsicherheiten von Messungen und deduktiven Schlüssen müssen identifiziert werden und Studien über deren Abschätzung werden von der wissenschaftlichen Literatur eingefordert. Basierend auf diesen Fakten haben wir uns die Forschungsfrage gestellt, ob ein datenzentrierter Ansatz erstellt werden kann, welcher die Daten- und Werkzeugsouveränität des forensischen Experten wahren und eine Voreingenommenheit bezüglich forensischer Werkzeuge verhindert werden kann und damit Verluste, Fehler und Unsicherheiten reduziert werden können.

Wir wenden unsere Forschung sowohl auf die digitale als auch auf die digitalisierte Forensik an. Die digitalisierte Forensik umfasst dabei die Anwendung von IT-Systemen für die Tatortforensik. Wir klammern dabei absichtlich die Assoziierungsketten und die Vorfallsrekonstruktion selbst aus und beschränken uns auf die in IT-Systemen enthaltenen digitalen Daten und deren Verwendung in forensischen Untersuchungen.

Wir verwenden in unserer Forschung die Begrifflichkeit der forensischen Untersuchung sehr breit gefasst. Sie reicht aus unserer Sicht vom technischen Support und Fehlersuche bis hin zum Einsatz für Gerichtsverhandlungen. Deshalb halten wir an den strengen Forderungen der umfassenden Nachvollziehbarkeit und den Einsatz wissenschaftlicher Methoden fest. Wir verwenden auch die Begrifflichkeit des IT-Systems ähnlich breit gefasst und beschränken uns nicht nur auf gewöhnliche PC-Systeme und Server. Wir verlangen von einem IT-System nur die Verarbeitung digitaler Daten, damit schließen wir beispielsweise auch eingebettete Systeme bis hin zum Cloudcomputing ein. Aus der Sicht der digitalen Forensik ergibt sich die Herausforderung aus dieser Heterogenität und Weite, eine gemeinsame Beschreibung der verarbeiteten Daten und der eingesetzten Funktionalitäten zum Einsatz in der Forensik zu finden.

Wir entwerfen eine fünfstufige Methodologie und stellen sie der Wissenschaft zur Verfügung, welche in unserem datenzentrischen Untersuchungsansatz (Data-Centric Examination Approach, DCEA) resultiert. Zunächst liefern wir eine formale Beschreibung von Verlusten, Fehlern und Unsicherheiten bezüglich der Daten in IT-Systemen. Die Basis bildet eine Modellierung anhand der Relationen zwischen den Daten, die jemals verfügbar waren, aller Daten aller forensischen Untersuchungen und fallspezifischer Daten eines gegebenen Vorfalls. Zweitens wenden wir ausgewählte Charakteristika des ISO/IEC 7498 Modells (allgemein bekannt als ISO/OSI Referenzmodell) auf Daten an, welche in IT-Systemen gespeichert, verarbeitet oder kommuniziert werden. Wir konstruieren Schichten von Daten mit Semantiken zur Unterstützung des forensischen Prozesses für die digitale Forensik (insgesamt 6) und für die digitalisierte Forensik (insgesamt 10). Drittens verwenden wir einen restklassenbasierten, hierarchischen Ansatz (konträr zu einer schichtenbasierten, sich nicht gegenseitig ausschließenden Beschreibung) zur Definition von Mengen von Methoden für den forensischen Prozess für die digitale Forensik (insgesamt 6) und für die digitalisierte Forensik (insgesamt 10). Dies schließt auch existierende Werkzeuge und Werkzeugsammlungen ein und verwendet ein existierendes Modell von Transferfunktionen zu deren Beschreibung. Zum Vierten verwenden wir ebenfalls einen restklassenbasierten, hierarchischen Ansatz zur Definition von Mengen von Untersuchungsschritten. Die Basis dafür bildet eine systematische Analyse existierender forensischer Prozessmodelle für die digitale Forensik. Wir definieren für die digitale Forensik 6 Schritte und adaptieren diese für die Verwendung in der digitalisierten Forensik, für welche dann ebenfalls 6 Schritte definiert werden. Abschließend verwenden wir die forensischen Datentypen, Mengen von Methoden und Mengen von Untersuchungsschritten für den forensischen Prozess, um eine qualitative Abschätzung von Verlusten, Fehlern und Unsicherheiten auf der Basis der Anwesenheit, Abwesenheit oder Diversität von forensischen Datentypen zu ermöglichen.

Wir testen unseren datenzentrischen Untersuchungsansatz DCEA an drei speziellen, aktiv erforschten Themengebieten aus digitaler und digitalisierter Forensik, bei denen sich die Beschreibung der forensischen Untersuchung besonders herausfordernd gestaltet. Dazu verwenden wir unsere veröffentlichten wissenschaftliche Beiträge für die Untersuchung von IT-Systemen zur Videoüberwachung für die digitale Forensik und Beiträge zur digitalisierten forensischen Daktyloskopie für die digitalisierte Forensik. Weiterhin führen wir eine bisher unveröffentlichte Forschungsarbeit zum Themenbereich der prozessor-kontrollierten Komponenten für die digitale Forensik durch. In unseren Experimenten für alle Anwendungsszenarien evaluieren wir insgesamt 31 forensische Methoden aus der Menge der Methoden für den forensischen Prozess für die digitale und digitalisierte Forensik. Wir durchlaufen die 6 Untersuchungsschritte und stellen Abschätzungen für 3 Fälle von Verlusten, für einen Fall eines Fehlers und 2 Fälle von Unsicherheiten fest.

Unsere Forschungen liefern weiterhin wesentliche Voraussetzungen für die Anwendung unseres datenzentrierten Untersuchungsansatzes DCEA. Diese umfassen die Analyse der IT-Systemlandschaft für die Abschätzung der vermutlich enthaltenen forensischen Datentypen, welche deshalb auch zumindest theoretisch zu sichern und auszuwerten sind. Weiterhin wird dadurch die Bestimmung der Systemgrenzen und der notwendigen Systeme zur nachfolgenden Untersuchung, Analyse und Dokumentation ermöglicht. Weiterhin ist eine kontextabhängige Definition der Menge von Untersuchungsschritten, der Mengen von Methoden und der forensischen Datentypen erforderlich. Zwingend erforderlich, wie die bisher unveröffentlichten Forschungsarbeiten zeigen, ist eine Festsetzung des Detailniveaus der forensischen Untersuchung, welche mit einer wohlgewählten Begründung einhergehen muss und die Grenzen für die Genauigkeit der qualitativen Einschätzung für Verluste, Fehler und Unsicherheiten für eine gegebene forensische Untersuchung setzt.

Durch unsere Forschung haben wir weitere Erkenntnisse geschaffen. Durch die Verwendung von Mengen von Untersuchungsschritten, von Mengen von Methoden für den forensischen Prozess und durch die forensischen Datenarten liefert der datenzentrische Untersuchungsansatz DCEA eine einheitliche Sprache. Diese einheitliche Sprache beschreibt Daten und deren Bearbeitung mittels Methoden des forensischen Prozesses und der Ergebnisdaten und deren Zusammensetzung, welche durch die Untersuchungsschritte räumlich und zeitlich zuordenbar werden.

Unser Ansatz kann, wenn die gleichen Definitionen für die forensischen Datenarten gewählt werden, für einen Intra-Untersuchungs-Vergleich für forensische Methoden verwendet werden. Dies könnte ein mögliches Kriterium für die Auswahl einer spezifischen Methode sein und kann für das Testen forensischer Werkzeuge verwendet werden.

Falls auch dieselben Definitionen für die Menge von Methoden des forensischen Prozesses und der Untersuchungsschritte gewählt werden, ermöglicht unser Ansatz auch den Inter-Untersuchungs-Vergleich. Ein derartiger Vergleich könnte unterstützend für eine vergleichende Evaluation des Reifegrades einer Untersuchung wirken und könnte für Fragestellungen zum Beweiswert hilfreich sein.

Die abschließende Fassung dieser Dissertationsschrift adressiert die zielführenden und inspirierenden Fragen und Kommentare meiner Betreuerin Prof. Dr.-Ing. Jana Dittmann sowie der Gutachter Prof. Eoghan Casey und Prof. Sabah Jassim aus den Gutachten und dem Kolloquium.

Legal notice / disclaimer:
The whole thesis is not devised to provide a legal contribution. The author and the contributors are no legal experts. All facts described herein are only used to derive technical and procedural requirements for the contribution of this thesis. Any reflections on legal issues are layman interpretations and are exclusively undertaken from a technical perspective.

Table of Contents

List of Figures

List of Tables

ACE-V ..... *Analysis Comparison Evaluation - Verification*
AFF ..... *Advanced Forensic File format*
AFIS ..... *Automated Fingerprint Identification System*
API ..... *Application Programming Interface*
ARP ..... *Address Resolution Protocol*
AS ..... *Autonomous System*
BGP ..... *Border Gateway Protocol*
BIOS ..... *Basic Input Output System*
BOT ..... *Beginning of Tape*
CAINE ..... *Computer Aided INvestigative Environment*
CAN ..... *Controller Area Network*
CC ..... *Methods for chain of custody maintenance*
CD ..... *Compact Disc*
CFA ..... *Cyber Forensic Assurance*
CFFTPM ..... *Computer Forensics Field Triage Process Model*
CFTT ..... *Computer Forensic Tool Testing Programme*
CL ..... *Methods for classification*
CMOS ..... *Complementary Metal Oxyde Semiconductor*
COI ..... *Community of Interest*
CPU ..... *Central Processing Unit, Central Processing Unit*
DA_{DF} ..... *Data Analysis for digital forensics*
DA_{SF} ..... *Data Analysis for digitised forensics*
DCEA ..... *Data-Centric Examination Approach*
DCI ..... *Methods for the digitalisation of contextual information*
DCO ..... *Drive Configuration Overlay*
DD ..... *Data types for digitised forensics*
DD_1 ..... *Raw sensor data*
DD_{10} ..... *Report data*
DD_2 ..... *Processed signal data*
DD_3 ..... *Contextual data*
DD_4 ..... *Parameter data*
DD_5 ..... *Trace characteristic feature data*
DD_6 ..... *Substrate characteristic feature data*
DD_7 ..... *Model data*
DD_8 ..... *Classification result data*
DD_9 ..... *Chain of custody data*
DEB ..... *Digital Evidence Bag*
DFIR ..... *Digital Forensics and Incident Response*
DFU ..... *Methods for data fusion*
DG_{DF} ..... *Data Gathering for digital forensics*
DG_{SF} ..... *Data Gathering for digitised forensics*
D_I ..... *Data containing information*
DI_{DF} ..... *Data Investigation for digital forensics*
D_{IF} ..... *Data containing forensically relevant information*
D_{IFC} ..... *Data containing case-specific, forensically relevant information*
DIKW ..... *Data Information Knowledge Wisdom*
D_{IR} ..... *Data of the investigation result*
DI_{SF} ..... *Data Investigation for digitised forensics*
D_{IT} ..... *Data of the investigation target*
DNA ..... *Deoxyribo Nucleic Acid*
DO_{DF} ..... *DOcumentation for digital forensics*
DOS ..... *Disk Operating System*
DO_{SF} ..... *DOcumentation for digitised forensics*

DPE ..... *Data Processing and Evaluation*
DPO ..... *Methods for digitalisation of physical trace objects*
DS_M ..... *Main memory data stream*
DS_N ..... *Network data stream*
DS_S ..... *Sensor data stream*
DS_T ..... *Mass storage data stream, Mass storage data stream*
DT ..... *Data types for digital forensics*
DT_1 ..... *Raw data*
DT_2 ..... *Hardware data*
DT_3 ..... *Details about data*
DT_4 ..... *Configuration data*
DT_5 ..... *Communication protocol data*
DT_6 ..... *Process data*
DT_7 ..... *Session data*
DT_8 ..... *User data*
DVD ..... *Digital Versatile Disc*
ECU ..... *Electronic Control Unit*
EEPROM ..... *Electrically Erasable Eletrically Programmable Read-Only Memory*
EIGRP ..... *Enhanced Gateway Routing Protocol*
ELF ..... *Executable and Linkable Format*
EMID ..... *Explicit means of intrusion detection*
EOD ..... *End of Data*
EOT ..... *End of Tape*
EPROM ..... *Electrically Programmable Read-Only Memory*
ERP ..... *Enterprise Resource Planning*
FAT ..... *File Allocation Table*
$f_{DF,SF}$ ..... *Transfer function in digital and digitised forensics*
FE ..... *Methods for feature extraction*
FIDEX ..... *Forensic Information Data EXchange*
FRE ..... *Federal Rules of Evidence*
FS ..... *File system*
GDPR ..... *General Data Protection Regulation*
GMT ..... *Greenwich Mean Time*
GUI ..... *Graphical User Interface*
HFS ..... *Hierarchical File System*
HPA ..... *Host Protect Area*
HW ..... *Hardware*
IBIS ..... *Integrated Ballistics Identification System*
IC ..... *Integrated Circuit*
IE ..... *Methods for image enhancement*
IEC ..... *Internation Eletrotechnical Commission*
IEP ..... *Information Exchange Packet*
IEPD ..... *Information Exchange Package Documentation*
ISO ..... *International Organization for Standardization*
ISP ..... *Internet Service Provider*
ITA ..... *IT Application*
LFTB ..... *Linux Transparent Forensic Bridge*
LR ..... *Likelihood Ratio*
MAC address ..... *Medium Access Control Address*
MAC times ..... *Modify Access Change or Modify Access Create (filesystem specific)*
MFD ..... *Multi Function Device*
MG ..... *Methods for model generation*
MMU ..... *Memory Management Unit*
MS ..... *Measurement support*

NAS ..... *Network Attached Storage*
NIC ..... *Network Interface Card*
NIEM ..... *National Information Exchange Model*
NIST ..... *National Institute for Standards and Technology*
NTP ..... *Network Time Protocol*
OEM ..... *Original Equipment Manufacturer*
OP_{DF} ..... *Operational Preparation for digital forensics*
OP_{SF} ..... *Operational Preparation for digitised forensics*
OS ..... *Operating system*
OSI ..... *Open Systems Interconnect*
OSPF ..... *Open Shortest Path First*
PA ..... *Methods for presentation and annotation of evidence*
PE ..... *Methods for parameter extraction*
PID ..... *Product IDentifier*
POSIX ..... *Portable Operating System Interface*
RAID ..... *Redundant Array of Inexpensive Disks*
RAM ..... *Random Access Memory*
RASP ..... *Random Access Stored Program*
RIP ..... *Routing Information Protocol*
SATA ..... *Serial AT Attachement*
SCSI ..... *Small Computer System Interface*
SEM ..... *Security Event Management*
SF ..... *Digitised forensics*
SMG ..... *Scaling of methods for evidence gathering*
SMO ..... *Sequential Minimal Optimization*
SNMP ..... *Simple Network Management Protocol*
SoC ..... *System on Chip*
SP ..... *Strategic Preparation*
SPAN ..... *Switched Port ANalyzer*
SP_{DF} ..... *Strategic Preparation for digital forensics*
SP_{SF} ..... *Strategic Preparation for digitised forensics*
SSC ..... *SCSI Stream Commands*
SSD ..... *Solid State Disk*
SW ..... *Software*
TBU ..... *Tape Backup Unit*
TCP/IP ..... *Transmission Control Protocol/Internet Protocol*
TOC ..... *Table of contents*
UDF ..... *Universal Disk Format*
UDP ..... *User Datagram Protocol*
UEFI ..... *Unified Extensible Firmware Interface*, *Universal Extensible Firmware Interface*
UFD ..... *USB Flash Disk*
URI ..... *Uniform Resource Identifier*
URL ..... *Uniform Resource Locator*
URN ..... *Unified Resource Name*
USB ..... *Universal Serial Bus*
UV ..... *ultraviolet*
VID ..... *Vendor ID*
VQI ..... *Visual Quality Indices*
WAN ..... *Wide Area Network*
WS ..... *Workstation*
XML ..... *Extensible Markup Language*

1. Motivation and introduction

In this chapter the main research topic is introduced and open challenges in the field of digital and digitised forensics are outlined to motivate the proposed data-centric investigation approach and the research into loss, error, and uncertainty. The research is motivated by a still unaddressed problem in the field of forensic sciences in general:

> "Few forensic science methods have developed adequate measures of the accuracy of inferences made by forensic scientists. All results for every forensic science method should indicate the uncertainty in the measurements that are made, and studies must be conducted that enable the estimation of those values." [CIN09, p. 184]

Further, in this section the main contribution of this thesis is summarised and its structure is introduced. For the whole thesis we work with the assumption that all IT systems involved in the process including the IT systems that form the target of the forensics examination and the IT systems used for the forensic examination are digital deterministic systems [Bur74, p. 295].

Some of the challenges outlined in the following Section 1.1 are the result of the fact that forensic science in general is a multi-disciplinary science with contributions from e.g. chemistry, physics, computer science, law etc., each with different definitions and models from their respective perspectives. Further, security research, of which a special branch is forensics, is special compared to other areas (e.g. software development) in so far as the whole research and its implementation have to happen with an adversarial environment in mind (also known as Anti-Forensics, see also [Gar07, p. 77]).

Other directions of research into digital forensics such as event reconstruction exist. They will not be covered in thesis. The reader is directed to [Gla04] and [Dew15, pp. 78-79] as a starting point. All of the above means to establish a sequence of events for an investigated incident share one important prerequisite - the existence of data from which the conclusions can be drawn. If this foundation is flawed, or important pieces of data is overlooked, the resulting theories regarding the most likely sequence of events are bound to be flawed. What is needed is a means to establish transparency of the data that is located, acquired, investigated, analysed and ultimately documented in the final report of the forensic examination. The Data-Centric Examination Approach (DCEA) introduced in this thesis in Section 4 attempts to add *transparency* with regards to the data that is processed in forensic examinations, especially in the light of loss, error and uncertainty, which is outlined in the following. We thank the reviewer Eoghan Casey for pointing us towards the traditional treatment of uncertainty to be measured by probability as described by the ENFSI Guideline for Evaluative Reporting in Forensic Science [Eur20, p. 6]. With DCEA we divert from this traditional definition of uncertainty by establishing a data-centric qualitative measure not for uncertainty but also loss and error.

1.1 Challenges in digital forensics and digitised forensics and resulting research question for this thesis

With the widespread use of IT systems, those systems became both targets for attacks and were used as tools to stage attacks and therefore are subject to forensic examinations (see [Cas11, pp. 9-12]). This gave rise to *digital forensics*, sometimes called computer forensics or forensic computing (see [SaJ07, p.1]) as a branch of forensic sciences. Although it shares many similarities with the traditional sciences, some properties are unique to the field of digital forensics. Digital forensics operates on *electronic evidence*, which is data and information of some investigative value that is stored on or transmitted by an electronic device, usually in digital form [New07, p.7].

Digital forensic examinations cover a wide range from embedded systems (e.g. in the automotive domain) over desktop IT systems to distributed client/server systems and to cloud computing. This thesis is based on the notion of *digital forensics*, as stated in [PBM08, pp. 114], that is often

used interchangeable with computer forensics, but digital forensics implies the inclusion of other than conventional desktop IT, such as network devices, mobile phones and other devices with embedded systems. For this thesis, this is further extended by adding networks of the aforementioned items. In this environment, potential examination targets for digital forensics span from very small embedded systems (e.g. a micro SD card) to a network of networks (employing a large number if interconnected IT systems acting as network nodes). A vast amount of data of very different types of data is being stored, processed and communicated, each of with different means of access and interpretation. From the viewpoint of digital forensics, this heterogeneity and vastness poses the challenge of finding a common description of processed data and processing functionality for use in forensics.

Another Challenge in digital forensics is that forensic tools are sometimes not comprehensively documented (closed source with its inner workings hidden from inspection). Further, some functionality of tools changes over time, some are not maintained (sometimes only used once to prove a hypothesis in scientific research) and some disappear altogether. Linking a forensic examination to a set of tools used is thus not forthcoming. This is particularly true if some examination needs to be revisited after some time and neither the hardware nor the software environment used in the examination is still available. The challenge is thus to find a comprehensive description of the forensic process that covers its properties and is independent of specific examination tools.

IT systems are increasingly used to assist the examinations of traditional forensic sciences. For the remainder of this thesis, the term *digitised forensics* is used to describe IT system assisted traditional forensic sciences. Such systems are used in ballistics examinations (e.g. see [FT14]), in micro-trace (fibre) examinations (e.g. see [Ver12, p.4]) and in forensic dactyloscopy (e.g. see [FBI14]), to name only a few branches of traditional forensic sciences. Those systems inherit properties of physical trace evidence and of digital evidence. The IT systems used to assist the traditional forensic sciences are, in principle, just as susceptible to IT-based attacks and incidents as IT systems used for other purposes. Existing solutions are typically offered by a niche industry, whose products are typically not comprehensively documented, especially the detailed inner workings, which form the intellectual property of respective manufacturers. Thus, in digitised forensics also there is a challenge to find a comprehensive description of the forensic process that covers its properties and is independent of specific examination tools.

We formulate a research question for both digital and digitised forensics as follows:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

After establishing the challenges in digital and digitised forensics and the research question for this thesis, in the following we will outline the methodology used to establish our findings outlined thereafter.

1.2 Methodology

We use a systematic analysis of existing approaches to establish a theoretical, data-centric model of the forensic process to derive qualitative measures of loss, error and uncertainty. To be able to do so, we define the term 'case-specific' as:

> Case-specific data:
>
> Data contained in an IT system (from single component up to networks of networks of systems employing a random access stored-program architecture [Har71, pp.234-240]) that can be a source of information leading to a successful event reconstruction of a suspected incident (on the target, victim and intermediate systems).

The methodology consists of the following steps:

M1. We describe loss, error and uncertainty regarding data contained in IT systems based on a modelling of the relationship of all data ever available, data used in all forensic investigations ever and the case-specific data for a given incident

M2. We apply the selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 621] to data stored, processed, communicated in IT systems to distinguish forensic data types for digital and digitised forensics. To achieve this, we construct layers of data by giving them semantics that support the forensic process. In further conjunction with the ISO/OSI model, we use a layering that is not mutual exclusive.

M3. We use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, pp. 3-4]. We further use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics.

M4. We use forensic data types, sets of methods and sets of examination steps to provide a qualitative estimation on loss, error and uncertainty.

With this methodology and addressing the challenges and the resulting research question in the next section present the contributions of this thesis.

1.3 Main contributions of this thesis

We believe in the absence of a sensible argument that would discourage us to pursue the course of research introduced and justified in the following. This Ph.D thesis delivers a formal, qualitative description of the forensic process by introducing the Data-Centric Examination Approach (DCEA), which is applicable to both digital and digitised forensics. Its main contributions are listed in the following:

C1. Formal description of loss, error and uncertainty using distinct sets of information contained in data (distinction into data containing information D_I, data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC} with only the latter solving specific incidents)

C2. Establishment of a data-centric view (data-driven) in digital and digitised forensics by applying a structured layering of data based on selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 621] in various IT systems (from embedded systems/IoT to data centers) to formulate forensic data types:

- C2.1 Adding semantics to data in the forensic context and establishment of 8 data types for digital forensics for selected use cases
- C2.2 Adding semantics to data in the forensic context and establishment of 10 data types for digitised forensics for selected use cases

C3. Establishment of a hierarchical mutual exclusive categorisation of methods of the forensic process:

- C3.1 Distinction into 6 distinct classes based on the likeliness of availability for digital forensics for selected use cases
- C3.2 Distinction into 10 classes based on the pipeline of the biometric process for a use case

C4. Establishment of 6 sets of examination steps for selected use cases:

- C4.1 Digital forensics process specific properties of the examination steps based on a systematic review of existing models and selection of best fitting for a data-centric approach
- C4.2 Digitised forensics process specific properties of the examination steps based on the application and adaption of the steps from digital forensics

C5. Case-specific qualification of loss, error and uncertainty and their representation based on forensic data types, sets of methods of the forensic process and sets of examination steps

Generally, we propose a *common language* to comparably describe different forensic examinations in a structured way, thereby allowing to systematically address issues of loss, error and uncertainty within procedures.

To show the universality and the utility of the approach, selected exemplary use cases from video surveillance, processor-controlled components and digitised forensic dactyloscopy are elaborated. Those were carefully chosen, since they are thought to be highly relevant but no known, universally agreed procedures to conduct the forensic examination in those application scenarios exist, yet.

The relevance for the *use case of video surveillance* stems from the fact that multimedia forensics, as outlined in [BFGK09, p.96], in an attempt to answer questions regarding media authenticity typically looks to unveil conspicuous traces of previous manipulations (manipulation detection scenario) and to infer knowledge about the source device (identification scenario). In this thesis, however, a different approach is described, which looks for difficult to avoid traces of multimedia content inside IT system. This approach is based on the *forensic data types*, *sets of methods for the forensic process* and *sets of examination steps* to extract multimedia and meta data in the light of *loss*, *error* and *uncertainty*, and thus outlining another way of authenticating multimedia content. This approach does not rely on sensor noise patterns and therefore avoids the some potential

attacks and remaining uncertainty attached to such identification scenarios [GFC11, p. 227]. Recovered multimedia content is tied to the IT system containing it, not to the sensor. This becomes particular relevant when in a surveillance scenario the sensor is easily changed but the system used to evaluate the incoming media content is fixed. The need to tie multimedia content to a device becomes even more relevant with ever more laptops, mobile phones and tablets being fitted with camera equipment.

As outlined in [Bee09, p. 24], non-standard computing environments still count as one of the largely unaddressed themes in digital forensics. Non-standard environments are also hidden away in the innards of common IT systems themselves. The relevance of the *use case of processor-controlled components* applies to such environments and is motivated by the fact that a lot of functionality of components of IT Systems is implemented using random access stored program (RASP, see [Har71, pp. 234-240]) logic. The functionality thus is dependent on the firmware inside those components. Since this firmware is often changeable using software only mechanisms, a real and present threat (see e.g. [Dom19]) e.g. for data hiding as a means of anti-forensics (see Section 2.6) exists. Although abstracted away using interface definitions and storage models, in fact, mass storage such as hard disk drives and USB thumb drives are not just containers for addressable sectors from 0 to maxSector() but IT systems in their own right with the prospect of (non-) maliciously altered operating environment (firmware). This, however, does not only apply to externally or internally connected mass storage. The concept of component with managed logic using firmware is applied to just about any component in an IT system with a reasonably complex functionality (e.g. graphics card, network controller, BIOS/UEFI etc.). This necessitates a defendable and justifiable decision about the depths of association (see Section 2.3) to be taken from operational preparation onwards (see Section 4.4.1.2) as to how deep the forensic examination has to look into the single components of IT systems of the investigation target to potentially detect such unauthorised alterations in firmware of the components. To conduct a forensic examination in this environment and in the context of *loss*, *error* and *uncertainty*, the data-centric forensic investigation approach is applied using the *forensic data types*, *sets of methods for the forensic process* and *sets of examination steps*, showing the versatility of the approach and its usability in this rather new area of digital forensics.

The relevance for the *use case of digitised forensic dactyloscopy* is motivated by the fact quite a lot of research is directed at the digitalisation of traditional crime scene forensics. Forensic dactyloscopy is a forensic discipline focusing on the comparative analysis and evaluation of fingermarks and fingerprints for individualisation purposes [Meu15, p. 734]. Digitised forensic dactyloscopy represents such an attempt to replace contact-based means of locating, acquiring, investigating and analysing fingerprint residue, in which introducing physical or chemical agents can hamper or forfeit further examinations of the same fingerprint residue [KHDV12, p. 1504]. New possibilities arise from the usage of contact-less fingerprint examinations, such as the ability to separate overlapping fingerprint patterns [QSS+12, p. 84361A] or estimating the age of a fingerprint residue [MBK+11, p. 41]. However, little research so far is directed at forensic processes in digitised forensic dactyloscopy, especially in the context of *loss*, *error*, and *uncertainty*. The usage of sensors to digitise aspects of the physical environment, potentially carrying a trace, and afterwards using only computational means to investigate and analyse this digital representation provides a challenge because of the demand for a comprehensive documentation for the whole examination, potentially for a non-expert [KHDV12, p. 1504]. In this scenario, to determine or verify a systematic course of action during a contact-less fingerprint examination, the data-centric examination approach introduced in this thesis is applied. *Forensic data types*, *sets of methods for the forensic process* and the *sets of examination steps* are adapted for the use in digitised forensic dactyloscopy, showing its versatility.

1.4 Thesis outline

The thesis structure is visualised with the following Figure 1. It comprises the sequential connection of each section to one another. The conceptual core is located in Section 5 and its application is described in Sections 6 to 8.

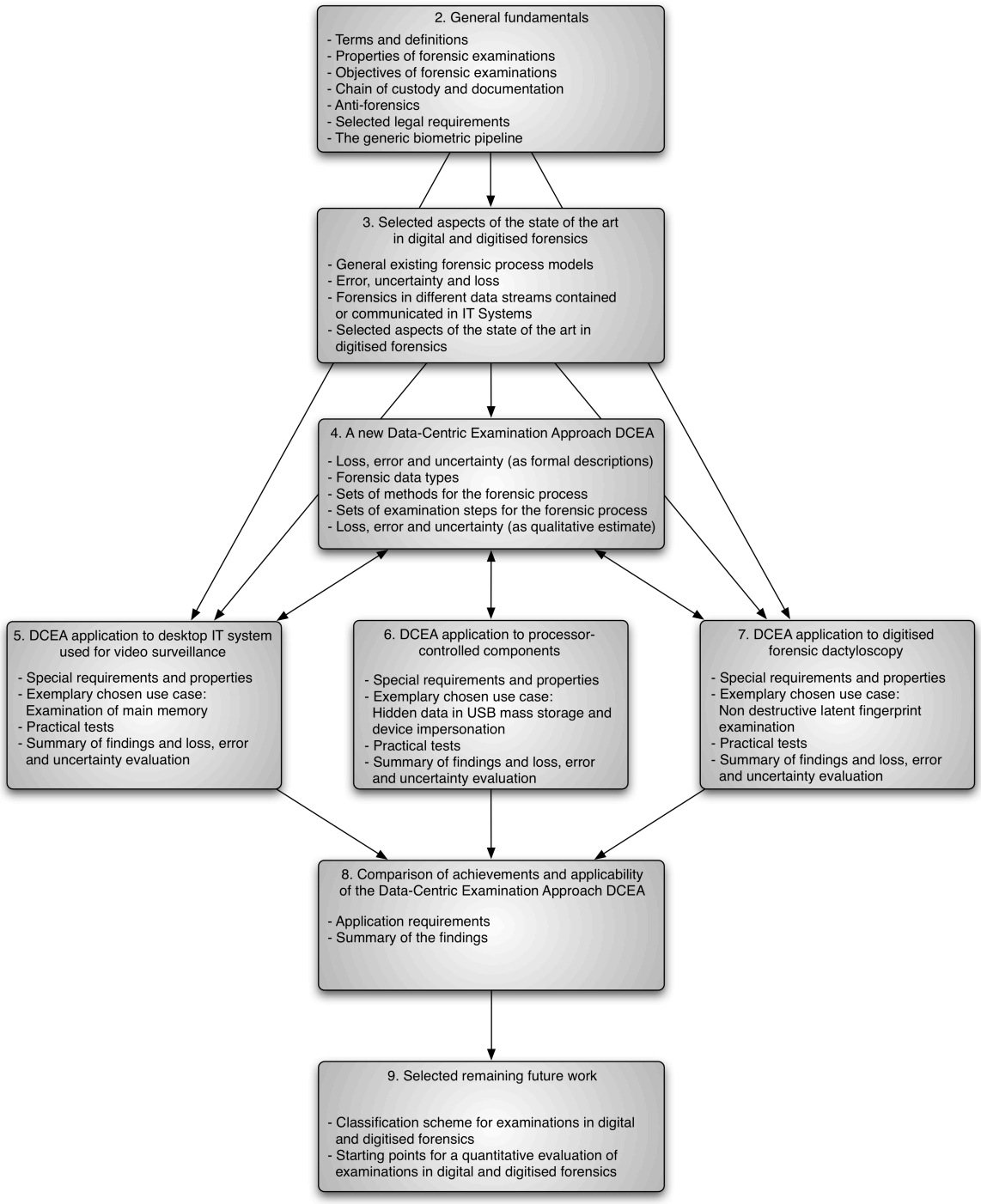

Figure 1: Thesis structure

Additionally to the sequential connections, the Section 3 containing fundamentals and Section 4 containing a selection of the state of the art directly support the application Sections 6 to 8. The findings of the application Sections 6 to 8 feed back information to the conceptual Section 5. The findings of the concept Section 4 and the application Sections 5 to 7 lead to the summary provided by Section 8. The thesis concludes with an outlook towards future work in Section 9.

2. General fundamentals

The contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Robert Altschaffel, Andreas Lang, Carsten Schulz, Claus Vielhauer (in descending order): [AKD09], [KLD08], [KHA+10]

In this Section the basics with regards to digital and digitised forensics are outlined. A wide variety of existing work and literature exists especially on the topic of digital forensics. We use selected literature to derive definitions and to show the state of the art without the goal of completeness. We provide definitions for selected key terms relevant for digital and digitised forensics.

Following that, the properties of forensic examinations in general are described in order to put the mindset of forensics across. This includes fundamental principles such as the divisibility of matter or the transfer of traits and matter, outlines the concepts of event reconstruction. The uniqueness of a forensic examinations and the necessity of minimising alterations to evidence, if not avoiding it altogether is motivated. Building on the reconstruction process during a forensic examination, the variety of potential objectives of such examinations is outlined.

Following that, the vital component of every forensic examination, the chain of custody is motivated and outlined. One potentially very important part in maintaining the chain of custody for electronic evidence suited for both digital and digitised forensics is presented in the shape of forensic evidence storage structures.

Taking into account that forensic examinations should always be thought of with the mindset of being in an adversarial environment and also to persecute IT security related incidents, the concepts of finding a common language for incidents and forensic investigations are motivated.

The problem of anti-forensics, i.e. means by which forensic examinations are hampered or derailed, is next motivated together with the presentation of a few exemplarily chosen attacker models.

Following that, selected important legal aspects that accompany forensic examinations are outlined together with a few principles directed at both the user of an IT system and the examiners.

As a support for digitised forensics in particular, the biometric pipeline for biometric user authentication is outlined next as it yields important insights on how biometric traits, one of which is the human fingerprint, can be processed by an autonomous system. Obviously those insights are very valuable when digitising the process of latent fingerprint examination in order to annotate and support the forensic expert. Finally, the existing knowledge necessary for the research challenge is summarised.

2.1 Terms and definitions

In this section, fundamental terms relevant to forensics and computer science are introduced for use throughout this thesis. Furthermore, the general problem of the anonymity of digital data is motivated, which impacts digital and digitised forensics.

2.1.1 Forensic examination, digital and digitised forensics, electronic evidence, post-mortem and live examinations

In [Pol08, p. 23], based on the findings of [InR01, pp. 15-17] a *forensic examination* is defined as "one or a series of investigative or legal questions, which are translated into scientific questions". It suggests a two-stage process with at first defining the legal questions and based on that the definition of scientific questions. However, in reality often the examiner asks the forensic examiner to extract any information from a given set of traces. Moreover, based on past experiences and knowledge, the forensic examiners themselves develop strategies based on the material pre-

sented to them. A forensic examination, at a very abstract level, aims to answer answering the following questions [AKD09, p. 54]:

- What has happened / is happening?
- Where has it happened / is it happening?
- When did it happen?
- Which way did it happen?
- What is the cause?

Traditional and digitised forensics also aim at answering the question:

- Who did it?

Answering the latter question especially for digital forensics is often difficult or outright impossible because of the anonymity of digital data. Some approaches, e.g. the one proposed in [VMC12, pp. 53-57], postulate to apply traditional or digitised forensics in some cases of digital forensic investigations, thus also trying to answer the question as to which individual is the originator of an incident. However, simply outside the scope of digital and digitised forensics is answering questions regarding the motivation for an individual to cause an incident.

For the remainder of this thesis, a *forensic examination* consists of localisation, acquisition, investigation and analysis. Inherent with of these steps (see also Section 4.4.1 of this thesis for a detailed description) is the process accompanying documentation and the dedicated creation of the final report. The terms acquisition and gathering are used interchangeably.

Notice the difference between the term *investigation* and *examination*. Often in literature those terms are used interchangeably. In this thesis those two terms denote different entities. For the remainder of this thesis the term *examination* describes the whole forensic process whereas the term *investigation* refers to a step of this examination (see Figure 2). However, when using direct citation, the original term is kept.

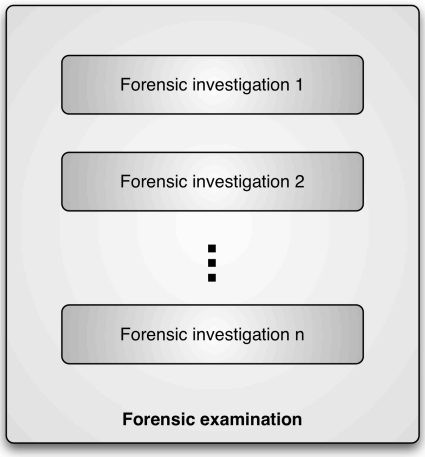

Figure 2: Forensic examination as a concatenation of forensic investigations

Digital forensics encompass, according to [Wol09, p.3], approaches and techniques for gathering and analysing traces of human and computer-generated activity in such a way that it is suitable in a court of law. Its objective is to perform a structured examination into past and ongoing occurrences of data processing and transmission whilst maintaining a documented chain of custody for the evidence, which can be reproduced unambiguously and validated by competent third parties [Wol09, p. 3]. As defined by [BFGK09, p. 92], digital forensics also includes multimedia forensics, since it also operates on electronic evidence. However, the goal of multimedia is to answer questions regarding media authenticity (see e.g. [Krae13, p. 1], which are of growing relevance and of particular interest in court, where consequential decisions might be based on evidence in the form of digital media [BFGK09, p. 95]. In the context of this thesis, multimedia forensics is acknowledged, as it provides potential means to ensure media authenticity (see also appendix Section 10.3), which is relevant also in digitised forensics. Although the details of multimedia

forensics are outside the scope of this thesis, the scenario presented in Section 5 tries to solve a problem from media forensics based on the data stored in a particular IT system.

In [AKD09, p. 54] a *broad definition of digital forensics* is given, which is more comprehensive as it includes the operator of an IT system as an influential factor of the degree of success and also deviates from a crime-centred view of digital forensics. For this thesis we extend this definition by main memory and conclude:

> *Digital forensics* is the strict methodological data analysis on storage devices main memory and in IT-networks for the purpose of solving incidents using the opportunities of strategic preparation from the viewpoint of the operator of an IT system.

We adopt this extended definition for the remainder of this thesis as it gives us more options for the strategic preparation and taking measures ahead of a specific incident.

No commonly accepted definition for digitised forensics exist in the literature, therefore for the remainder of this thesis the following will be used:

> *Digitised Forensics* cover the localisation and digitalisation of physical trace evidence and the subsequent, exclusively digital, examination in support of a forensic expert for usage in the court of law, whilst maintaining a physical and a digital chain of custody and providing means to reproduce the findings by a third party. Computational Forensics as described in [Sri10, pp. 39-43] has very similar goals and methods as digitised forensics. Both terms depict a relatively new branch of forensic sciences (see also [Sri10, p. 40]).

In digital and digitised forensics, the objective is to locate, acquire, investigate and analyse *electronic evidence*. Thanks to Eoghan Casey from his review for the remainder of this thesis we use the consensus standard definition for electronic evidence as information of probative value that is stored or transmitted in binary form [CBT20, p. 1] citing ASTM E2916-13.

According to an alternative definition from [New07, p.7], electronic evidence is data and information of some investigative value that are stored on or transmitted by an electronic device, usually in digital form. Electronic evidence, typically, is latent evidence, i.e. it is difficult to view and needs to be rendered visible during the stages of the forensic examination.

Others, such as [CaS05, p. 1] go even further and argue that *digital evidence of an incident* is digital data that contain reliable information that support or refute a hypothesis about the incident being investigated. In [Pol01, p. D4.89], digital evidence is defined as information of probative values stored or transmitted in digital form. For the remainder of this thesis, the terms electronic and digital evidence are used interchangeably since it is only concerned with electronic information that is contained in digital data as opposed to analogous information (e.g. cassette tape audio/video recording). In Section 2.7.2 of this thesis a more detailed discussion on the evidentiary value of electronic/digital evidence for digital and digitised forensics is undertaken.

Particular in digital forensics, a distinction between *post-mortem examinations* and *live examinations* can be drawn [CVD+10, p. 307]. Post-mortem examinations are conducted typically after an incident has passed and require the examiner to shut down the machine to inspect the contents of mass-storage devices and identify artefacts of interest. However, this process breaks network connections and unmounts encrypted disks causing significant loss of potential evidence and possible disruption of critical systems. Live examinations are conducted typically during an ongoing incident and can allow an examiner to inspect the state of a running IT system without disruption. However existing tools can overwrite evidence present in memory or alter the contents of the disk causing forensic taint, which lowers the integrity of the evidence.

2.1.2 Data and information, syntax and semantics, signal

In digital and digitised forensics the terms data and information play a vital role. Out of the many definitions available, those from [Row07, p. 166] citing [Ack89] fit the needs of this thesis.

According [Row07, p. 166] citing [Ack89], *data* is defined as symbols that represent properties of objects, events and their environment. They are the products of observation. But are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural.

Information as described in [Row07, p. 166] citing [Ack89] is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate, store, retrieve and process data. Information is inferred from data.

Derived from linguistics and used in computer science, syntax and semantics can play an important role in digital forensics and digitised forensics as it allows looking at different aspects of a given object.

In [Cho65, p.16] as syntactical component, which we will refer to as *syntax*, is specified as an infinite set of abstract formal items, each of which incorporates all information relevant to a single interpretation of a particular sentence. A sentence is represented in the context of this thesis as a set of digital objects.

In [Nie95, p.1] (formal) *semantics* is specified as being concerned with rigorously specifying the meaning, or behaviour, of programs, pieces of hardware etc. This is particularly relevant in digital and digitised forensics as it gives meaning to data contained in digital objects, which can be retrieved by software or by human examiners. Semantics is a precondition for the reconstruction process introduced in the following section 2.2.

Particularly important for digitised forensics are *signals*. According to [Lah98, p. 51], a signal is a set of information or data that is a function of an independent variable (e.g. time, space). Signals are usually processed by systems that modify them to extract additional information from them.

2.2 Properties of forensic examinations

Forensic examinations rely on the formulation of investigative questions and using scientific methods to answer those questions. On a very abstract view, according to [InR01, pp. 77-79], this involves *principles* and *processes*. Therefore, in the following the two principles of transfer of matter and traits as well as the divisibly of matter are outlined and put in the context of digital and digitised forensics. Afterwards, the processes of identification, classification, individualisation, association and reconstruction and their connections to digital and digitised forensics are outlined. Note, that the term classification in the context of the work in [InR01, pp. 116-117] has a different meaning compared to its use in biometrics and pattern recognition outlined in Section 2.8.

2.2.1 Principles: Transfer of traits and matter, divisibility of matter

One approach to develop the investigative questions is the result of the application of two principles *transfer of traits and matter and divisibility of matter* (see [InR01, p. 77]). The first principle of *transfer of traits and matter* implies that in the physical world just about everything can become a trace, as soon as it relates to an examined case. This is important since virtually impossible to conduct any action without leaving any trace either by introducing material to the scene or by taking in material from the scene (see [InR01, p.84] citing [Loc20]). It was first postulated by Locard in 1920 (see [InR01, p.18] citing [Loc20]). In digital forensics the principle of transfer of matter does not apply, since in the digital domain no physical matter is transferred.

However, the principle of *transfer of traits* is heavily used, as digital objects interact with each other and influence each other. In digitised forensics the principle of transfer of matter applies since physical trace evidence could have been subjected to matter transfer before digitalisation.

The second principle of the *divisibility of matter* states that matter can be divided into smaller component parts, given that sufficient force is applied. This leads to three relevant conclusions [InR01, p. 87]:

- Some characteristics retained by the smaller pieces are unique to the original item or to the division process. These traits are useful for individualising all pieces to the original item;
- Some characteristics retained by the smaller pieces are common to the original item as well as to other items of similar manufacture. We rely on these traits to classify the item;
- Some characteristics of the original item will be lost or changed during or after the moment of division and subsequent dispersal; this confounds the attempt to infer a common source.

The principle of the divisibility of matter is relevant for both traditional and also for digitised forensics since the physical trace evidence could contain matter that has been divided before the digitalisation. The following Figure 3 illustrates the connections between the principles and the forensic disciplines based on the observations above.

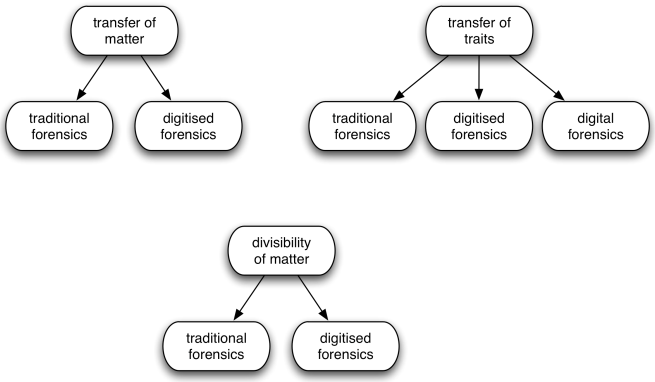

Figure 3: Transfer of matter/traits and divisibility of matter in forensic sciences as it applies to traditional, digital and digitised forensics (based on observations about the transfer of matter and transfer of traits in [InR01, pp. 98-99])

Using those two principles, as suggested in [InR01, p. 77], in the following Section 2.2.2 five processes are outlined.

2.2.2 General processes in forensic examinations

In traditional forensic sciences and with modifications also for digitised forensics and for digital forensics the following *processes* are important for the planning of forensic examinations [Pol08, pp. 20-23]:

- *Identification:* Concept of the physiochemical nature of trace evidence in order to be able accurately describe an item or its composition, also in digital forensics describing evidence in terms of its context physically (e.g. a brand of harddrive), structurally (e.g. number of cylinders, heads), logically (e.g. FAT partition), per location (e.g. directory) or per content (e.g. spreadsheet, email),
- *Classification:* Classification allows for attempting to determine a common origin of trace evidence, also in digital forensics describing characteristics (e.g. file systems, partitions, files)
- *Individualisation:* individualisation uses a set of characteristics to uniquely identify a specific specimen, also in digital forensics describing individual specifications (e.g. cryptographic hash sum of a file)
- *Association:* As defined in [InR01, p. 168], this allows for an inference of contact between the source of the evidence and a target, also in digital forensics describing links between digital objects (e.g. presence of specific files linking them to the perpetrator and/or victim IT system)

- *Reconstruction:* As defined in [InR01, p. 79], this allows for an ordering of associations in space and time, which also applies to digital forensics.

The following Figure 4 summarises and structures those processes (based on [InR01, pp. 77-79]).

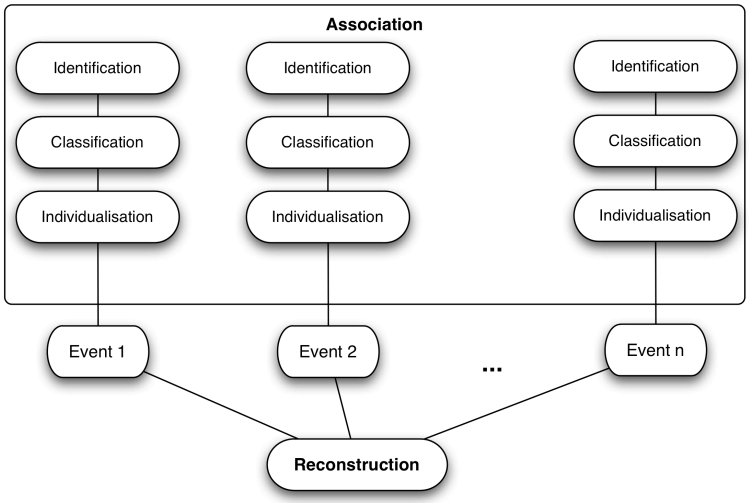

Figure 4: Structuring based on processes [InR01 pp. 77-79] in a forensic examination

Forensic examinations often involve the *reconstruction* of *events*. Thanks to the reviewer comment from Eoghan Casey for reconstruction we adopt the summary from [PCJ+18, p.7] with reconstructions involving temporal, relational and functional analysis.
In this context an event is that objects came into contact with each other. For the reconstruction of events they need typically multiple *associations*, which can be made using the *Identification*, *Classification* and *Individualisation* of objects.
We thank the reviewer Eoghan Casey for pointing out that event reconstruction is not the solely purpose of forensic examinations. Forensic examinations can also be only concerned with comparisons such as classification and identification whilst others are concerned with authentication (see also Section 2.3)
The appendix Section 10.1.1 contains further reflections regarding this topic in the legal context.

2.2.3 Uniqueness of forensic examinations

Another important property of forensic examinations is that each forensic examination needs to be treated as *unique.* One of the most common mistakes is to apply case-specific knowledge to other, similar-looking cases as stated in [Cas11, p. 53]: "Even experienced examiners are prone to forming such preconceived theories because they are inclined to approach a case in the same way as they have in approached past cases, knowing that their previous work was upheld". As stated in [Pol08, p. 23], this often results in examinations designed based on what could be done instead of on the specific information that could be located. In the legislation of some countries this "fishing for evidence" is prohibited.

2.2.4 Alterations to evidence (minimisation and explanation)

One of the most prominent demands placed onto forensic examination is to *minimise changes* to the examined objects or to avoid them altogether, thus keeping the *fidelity* of trace evidence [SaG09, p. 120]. If changes cannot be avoided, at least they have to be properly explained (e.g. by known characteristics of the used forensic tools and techniques). This is particularly relevant in traditional forensics (e.g. contact-based forensic dactyloscopy) and live examinations in digital forensics (see Section 3.3.1.2).

Although IT systems are used in digital and digitised forensics, it is a basic principle [RYG05, p.6] that the critical decisions are left to human experts, no matter how sophisticated the automated examination systems and solutions might get (see also Section 2.7).

2.3 Objectives of forensic examinations

The most obvious and commonly named objective of a forensic examination is to reconstruct events of an incident in order to identify the person(s) responsible, typically to bring about some sort of punishment after being found guilty in the court of law. “The aim of (digital) forensic analysis remains the same – to clarify events of the incident and, ultimately, identify its perpetrators” [Gla04, p. 1].

However, the underlying principles of forensic sciences, most importantly the demand for a fully comprehensible conduct of the examination, the maintenance of a chain of custody (see Section 2.4) and the usage of commonly accepted scientific methods (see Section 2.7), as well as maintaining the security aspects of integrity, authenticity and confidentiality can in fact be applied to a whole range of activities.

One such activity is, for instance, the intrusion detection with the objective of closing security vulnerabilities to prevent similar intrusions from happening/succeeding again [KH02, p.4]. But also internal corporate examinations without the aim of involving the legal system fall into this category. Furthermore this also typically applies to the case of finding the non-malicious root causes of malfunctioning IT systems, which profit from the application of forensic principles. Such application scenarios can include support cases such as malfunctioning hard- and software or the operating errors [KHA+10, p. 101]. This fact is also underlined in [KCG+12, p. 2-1 - 2-2], where the need for forensics points towards:

- Operational troubleshooting (e.g. misconfigured networks of physical or virtual hosts, resolving functional problems with IT applications),
- Log Monitoring (e.g. assisting incident handling, identifying policy violations),
- Data recovery (both malicious and non-malicious causes),
- Data acquisition (retired or redeployed IT systems),
- Due diligence/Due Compliance (sensitive information protection, audits).

Adhering to the aforementioned requirements placed on forensic examinations in all cases, following a model of the forensic process such as the one introduced in Section 4, allows a very flexible response depending on the results gained. As stated in [KH02, p.3], each examination should be treated as if it would end up in court. This way, suspected non-malicious incidents according to the symptoms, can be used in legal proceedings, should it become evident that the symptom is indeed the result of malicious activities.

Depending on the objectives of the examination, the chain of examination processes according to [InR01, p. 115] of Association (Identification, Classification, Individualisation) does not necessarily have to be fully completed (see Figure 5).

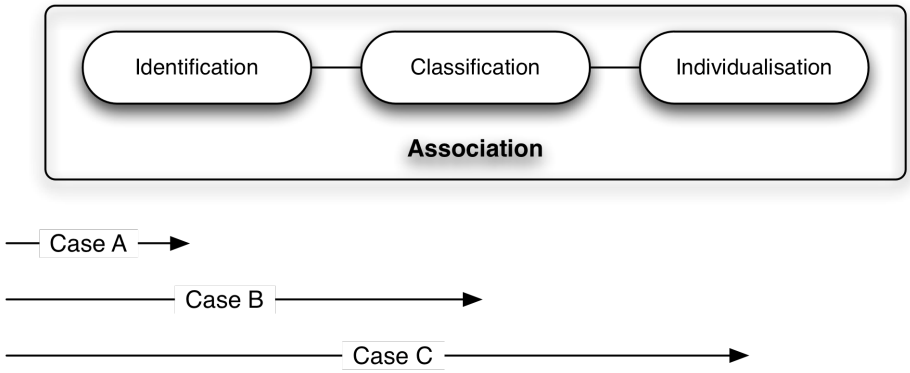

Figure 5: Case-specific depth of the association depending on the objectives of a forensic examination based on identification, classification and individualisation [InR01, p. 115]

The depth of the association depends heavily on the objectives of the forensic examination. This is demonstrated in the following using exemplarily chosen cases. If, for instance, in a non-malicious *Case A* in digital forensics a misconfigured dhcp server assigns the same IP address to two different IT systems, it is sufficient to only finish the Identification process (ARP table comparison after recording the IP traffic). However, if in a second *Case B* from traditional forensics both the Identification and the Classification result in a finding that a person possessed drugs (white powder, person showing symptoms of intoxication) and those could be classified Class A drugs (Cocaine), this classification based on possession alone is sufficient for legal proceedings in a lot of countries. If, in a third *Case C* in digitised forensics, microscopic paint fragments are found on the victim's clothing in a suspected hit-and-run accident, those paint fragments need to pass positive micro trace identification during the Individualisation (fragments that fit to other samples collected from a suspected vehicle) after the Identification (paint micro traces) and Classification (paint typically used in automotive production) took place.

2.4 Chain of custody and documentation

Documenting the entire forensic examination is crucial and decides over the admission of the examination results of evidence. As outlined by [Mac11, p. 10-4] the documentation establishes transparency of all activities and data generated. It needs to support the reported conclusions and must contain sufficient detail.

One important part of the documentation is the chain of custody. As stated in [New07, p. 6] a "chain of custody is the route the evidence takes from the initial possession until final disposition. This documentation process is one of the most critical of the investigation." A chain of custody is centred around the items of evidence and according to [KH02, p. 8] the following questions need to be answered in detail:

1. Who collected the evidence,
2. How and where was the evidence collected,
3. Who took possession of the evidence,
4. How was the evidence stored and protected in storage,
5. Who took the evidence out of storage and why.

As stated in [KH02, p. 13] "the chain of custody and other evidentiary handing rules assure the jury that no unanticipated or introduced changes occur and that it is reasonable to extrapolate from the point of collection back the time of the incident." This amounts to protecting the integrity of the evidence. By further documenting answers to how and where was the evidence collected (2), the security aspect of authenticity is addressed (assuming that appropriate organisational or technical measures are taken). Answering the questions who took possession of the evidence (2) and who took the evidence out of storage and why (5), also the security aspect of non-repudiation is addressed (again, assuming that appropriate organisational or technical measures are taken).

In traditional forensics, different types of custody can occur, as outlined on the example of latent fingerprint examination in [Mac11, p. 10-4]. Here, three different types of custody are named as *primary custody*, *secondary custody* and *tertiary custody*. The differentiation is drawn according to when the examiner receives custody of the items of evidence. *Primary custody* is established when the examiner initiates the chain of custody for both the item of evidence and the recovered fingerprints and maintains it from the discovery of the evidence through the examination. This covers the situation of an examiner responding to a crime scene, locating and acquiring the items of evidence, and transports the items to the laboratory where he investigates and analyses the fingerprints off the item of evidence. *Secondary custody* is established when the examiner receives the items of evidence secured by other personnel, for instance a crime scene analyst. Here the examiner acquires, investigates and analyses the fingerprints off the item of evidence. In this case the examiner initiates the chain of custody for the recovered fingerprints but not for the item of evidence. *Tertiary custody* is established when the examiner receives fingerprints for analysis

with the localisation, acquisition and investigation of items of evidence being carried out by other personnel such as a crime scene analyst. Here the examiner maintains the chain of custody for the fingerprint. Those types of custody also apply for digital and digitised forensics, which this thesis is concerned with.

In digital forensics the item of evidence always includes data items and often the physical items that contain the data. For both types of items a chain of custody needs to be maintained. However, in some cases of digital forensics the chain of custody can comprise only data with no physical items seized (e.g. a network traffic based examination). Digitised forensics always involves physical items and data items as the digital representation of acquired signals from the physical object. Here a separate chain of custody for both types needs to be initiated and maintained. This is particularly relevant as outlined in Section 4.4.2.2 of this thesis.

We thank reviewer Eoghan Casey for pointing out that the chain of custody is only a part of the concept of provenance (see e.g. [LeL09, p. S49] and [CBB15, p. S106), which also includes the methods used to process data. We address this aspect in Section 4.3 of this thesis in detail.

The important problem of documenting forensic examination, which includes both the chain of custody and aspects of provenance, is part of the Data-Centric Examination Approach DCEA introduced in this thesis in Section 4.

2.5 Incident-Taxonomy and Forensic Examination Taxonomy

One important goal of digital forensics is the examination of IT security related incidents. Also, should both an examination in digital or digitised forensics become attacked itself; this incident needs to be examined as well. One important precondition is to find a common language to describe an IT security related incident. This mirrors the demand in forensics for a description for the evidence recognition process in a general way [Car06, p. 1]. In the following, proposals addressing the description of IT security related incidents and forensic examinations are outlined.

2.5.1 The incident taxonomy

In [HoL12, pp. 1-32] the construct of a *taxonomy* is chosen for this description. According to [HoL12, p. 2], a taxonomy has the following properties of being:

- mutually exclusive - classifying in one category excludes all others because categories do not overlap,
- exhaustive - taken together, the categories include all possibilities,
- unambiguous - clear and precise so that classification is not uncertain, regardless of who is classifying,
- repeatable - repeated applications result in the same classification, regardless of who is classifying,
- accepted - logical and intuitive so that categories could become generally approved,
- useful - could be used to gain insight into the field of inquiry.

Further, the incident taxonomy is designed using a minimum set of high-level terms and comes along with a structure that indicates their relationship in order to describe security incident and vulnerability information [Hol12, p. 1]. It enables the integration of lower-level terms of specific incidents into the common language. The incident taxonomy has both a horizontal and a vertical component (see Figure 6).

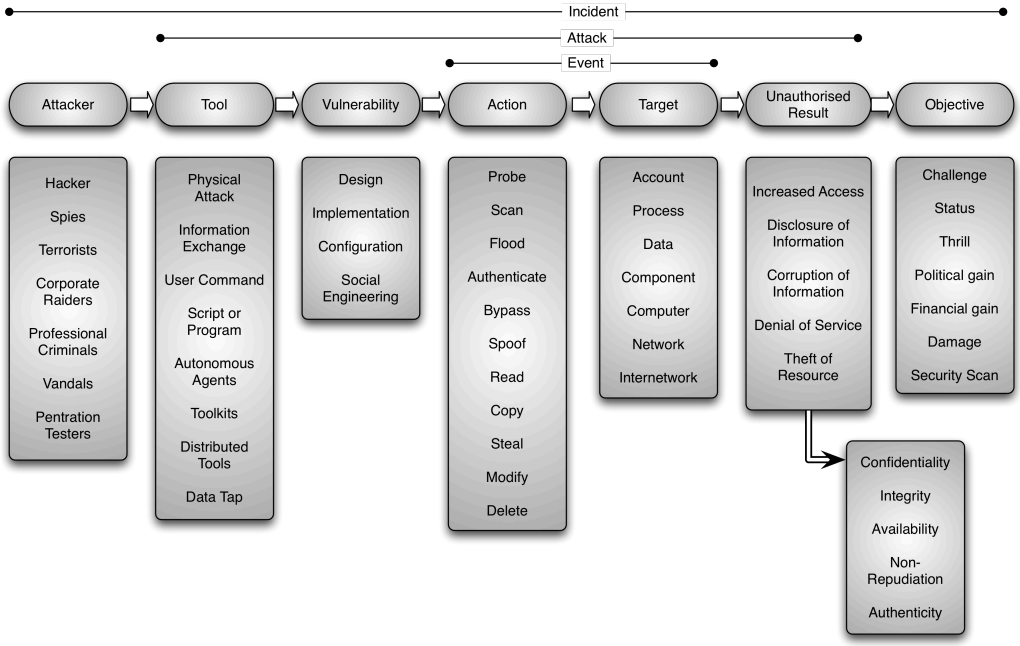

Figure 6: Incident taxonomy [HoL12, p. 16], enhanced by adding penetration testers, social engineering, security scans and integrating IT security aspects [KLD08, p. 414]

The horizontal component describes the relationship between:

- *event* - a discrete change of state or status of a system or device, result from actions that are directed against specific targets,
- *attack* - a series of steps taken by an attacker to achieve an unauthorised result,
- *incident* - a group of attacks that can be distinguished from other attacks because of the distinctiveness of the attackers, attacks, objectives, sites, and timing

by using seven main categories. Those categories are [Ho12]:

1. Attacker - an individual who attempts one or more attacks in order to achieve an objective,
2. Tool - a means of exploiting a computer or network vulnerability,
3. Vulnerability - a weakness in a system allowing unauthorised action,
4. Action - a step taken by a user or process in order to achieve a result,
5. Target - a computer or network logical entity (account, process, or data) or physical entity (component, computer, network or internetwork),
6. Unauthorised Result - an unauthorised consequence of an event,
7. Objective - the purpose or end goal of an incident.

In the horizontal component view of the incident taxonomy, an *event* comprises of an action directed towards a target. When observing more details, the *attack* becomes visible, comprising of actions directed against targets using a tool and exploiting a vulnerability, yielding a result. By getting even more insight, the whole *incident* is visible, comprising of actions directed against targets using a tool and exploiting a vulnerability, yielding a result, initiated by an attacker in order to achieve an objective.

The vertical component view contains the items inside the categories for each category, respectively. By using the items from the categories and formulating events, attacks and incidents, a specific IT security incident can be described using the abstract taxonomy terms.

As stated in [HoL12, p. 1], those lists of items should be extended to adapt the incident taxonomy to changes. In [KLD08, pp. 415-416] a number of extensions is proposed. Those entail the adding the item penetration tester to the attackers, to account for security evaluations. To correspond with

this, the objective of security scan is added. Further, to account for a lot of types of incidents, where no security breaches on the respective IT systems but instead the weakness of the human operator is exploited, the vulnerability offered by social engineering is added. Further, to detect violations of security policies, the detection of the violation of security aspects is essential. Therefore, the security aspects were added, not as a new category but a means to further describe the results category, serving as a reminder to update security policies in the light of events.

2.5.2 The forensic examination taxonomy

The notion of finding a common language and using a taxonomy approach is also applied to digital forensics in [AKD09, pp. 54-68]. The motivation stems from the fact that here too, a common language is necessary (e.g. for exchanging examination strategies in similar cases). Broadly speaking, the categories need a new horizontal arrangement and new categories and items, whilst disregarding a number of categories and items (see Figure 7). This is due to the nature of forensic examinations. Pinning down an attacker (as a physical individual) and his objectives are typically beyond the capabilities of digital forensics.
In total three new categories are added:

- Violated security aspects - failures of the security policies
- Timeline - differentiation between post-mortem and live digital forensics
- Origin - distinction between a malicious or a non-malicious incident

In the existing category of vulnerability, the item of human behaviour is added, reflecting non-malicious activities by users of a system, resulting in an incident.

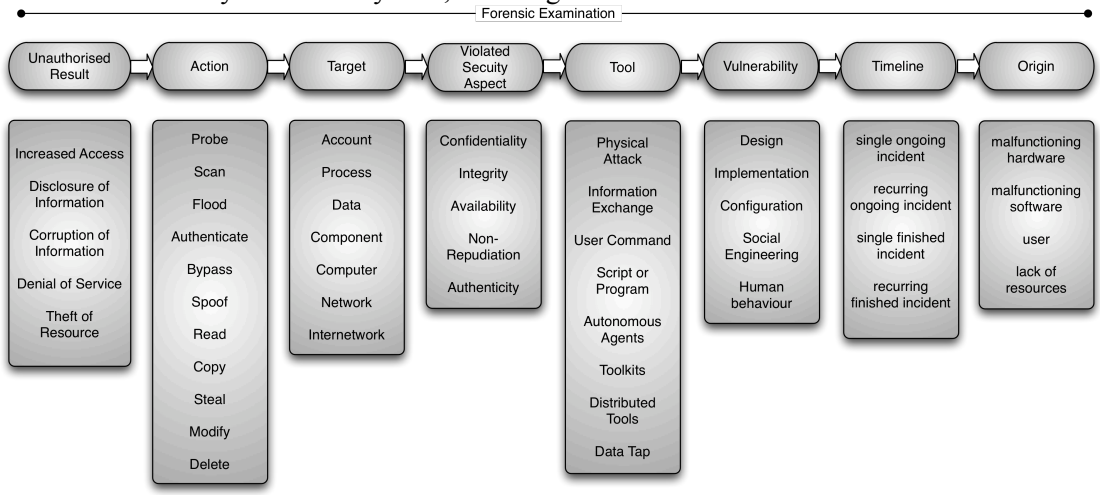

Figure 7: Forensic examination taxonomy, modified from [AKD09, p. 65]

For the general questions regarding a forensic examination from Section 2.1.1, in [AKD09, p. 65] a mapping of those questions onto the categories from the incident taxonomy is proposed:

- Where has it happened/is it happening? - This is described by the existing category target,
- Which way did it happen? - This is described by the existing category action,
- What is the cause? - This is described by the extended category vulnerability,
- When did it happen? - This is described by the newly proposed category timeline,
- What has happened/is happening? - This is described by the moving of the security aspects into a separate category violated security aspects.

The forensic examination taxonomy reflects in its horizontal arrangement of categories an abstraction of the knowledge gathered during an examination in digital forensics. In Section 4.4 of this thesis, the underlying examination steps are modelled.

2.6 Anti-Forensics

IT security, of which forensics is a special branch, is different from other aspects of computing due to the constant potential presence of an adversary (see [Gar07], pp. 77-84). Hence, identifying and addressing security vulnerabilities requires a different mindset compared to traditional engineering due to the fact that it takes place in an adversarial environment also applicable to digital and digitised forensics (see additional reflections in appendix Section 10.2 about attacker modelling). Methods of the forensic process can be deliberately confused. That branch of research is also called anti-forensics. As stated in [Gar07, p. 77], anti-forensics is a growing collection of tools and techniques that frustrate forensic tools, examinations and examiners. According to [Gar07, p. 77] citing [LiB06], anti-forensics has four primary goals:

- avoiding detection that some kind of event has taken place,
- disrupting the collection of information,
- increasing the time that an examiner needs to spend on a case, casting doubt on a forensic report or testimony.

In addition to that, other goals are imaginable, as pointed out by [Gar07, p. 77]:

- forcing the forensic tool to reveal its presence,
- subverting the forensic tool (e.g. using the forensic tool itself to attack the organisation in which it is running),
- mounting a direct attack against the forensic examiner (e.g. discovering and disconnecting the examiner's network or mounting physical attacks against the examiner's workplace),
- leaving no evidence that an anti-forensic tool has been run.

Anti-forensics directed at digital forensics can include the (selective) overwriting of content of mass-storage devices, the employment of cryptography or steganography or other data hiding approaches, malicious code techniques such as memory injection, the usage of live media booted off a secondary media and detection of forensic tools and exploiting known vulnerabilities therein, just to name a few examples (see e.g. [Gar07, pp. 77-84] for more details).

Although somewhat controversial, researchers in anti-forensic claim to challenge the so-called "presumption of reliability" and that they are not creating vulnerabilities, they just identifying them (see [Gar07, p. 77]). In doing so, they demonstrate research potential for the continuous improvement of forensic tools and methods.

After elaborating what the notion of forensic computing in an adversarial could entail, in the following selected legal and data protection requirements are described.

2.7 Selected legal requirements

Disclaimer: The whole thesis is not devised to provide a legal contribution. The author and the contributors are no legal experts. All facts described herein are only used to derive technical and procedural requirements for the contribution of this thesis.

The statements about legal requirement made in this section are highly dependent on the jurisdiction and the existing laws of a particular country. Further it has to be stated that the author of this thesis has no background in law and legal proceedings. All statements, although thoroughly researched and referenced, cannot and are not intended to replace expert knowledge from a law professional. However, this exemplary selection of legal requirements is intended to give the reader a particular mindset that is necessary when dealing with forensic examinations (in addition to the mental grammar of operating in an adversarial environment as motivated in Sections 2.5 and 2.6). A further selection of additional, connected information is provided in the appendix in Section 10.1. As a general principle forensic examiners have to look for both inculpatory evidence (that which supports a given theory) and exculpatory evidence (that which contradicts a given theory as well as for evidence of tampering [Car03a, p. 2]. In [Car03, p. 8] this amounts to comprehensiveness.

2.7.1 “Null” hypothesis and Burden of proof

One important aspect for all forensic examinations in a lot of countries is the basic principle of “Innocent Until Proven Guilty”. The Burden of Proof [Dar10, p. 58] in criminal prosecutions lies with the prosecutor. For Under U.S. law, one is considered innocent until proven guilty. In research terms this would be equivalent to having to reject the “Null” hypothesis to prove guilt. In effect, we are proving that the defendant is “not innocent”. The defendant does not need to prove the “Null” hypothesis - however, the prosecutor needs to prove that the “Null” hypothesis needs to be rejected.

2.7.2 Evidentiary Value

One important aspect regarding the evidentiary value of digital data used as electronic evidence is the trust placed on the digital object based on the security aspects, especially with regards to authenticity and integrity. Adding to that, the appendix Section 10.1.1 also provides additional information about the need to validate the evidence. The challenge is that digital data, in general, can easily be manipulated/forged.

In digital forensics, the value of forensic evidence, according to [Car05, pp. 18-19], depends on it being *essential data* or *non-essential data*. Essential data describes data that is necessary for a (sub-) system to serve its intended purpose, e.g. for a file contained in a file system to be useful, the link between the file name and its content needs to be correct. In other words, the manipulation of essential data takes a considerably higher effort from the malicious user - the (temporary) correct use of essential data and its subsequent alteration is bound to leave traces. Non-essential data on the other hand only adds non-essential functionality, e.g. in a file system the alteration or corruption of MAC times would not disallow to access the file itself.

Since the functionality of the system for a given purpose is also relevant of a malicious user, essential data is of higher value as evidence compared to non-essential data, given the precondition that also the malicious user needs a functional (sub-) system. Non-essential data may be of relevance in forensic examinations; however, extra effort has to be invested form the position of the forensic examiner to show that this piece of data was not altered. Simply put, essential data have a higher level of trust assigned to it. However, careful consideration has to be taken, what type of data, and in what context, is considered essential data.

This idea about essential and non-essential data is elaborated further in [Dew12, pp.40-41]. Here a distinction is made between *technically unavoidable traces* (in this context this equals data), which occur inevitably and thus cannot be avoided by simple changes to the respective system. Technically avoidable traces (i.e. data) are traces that are created for their own sake. Again, a higher evidentiary value is to be placed on technically unavoidable traces due to the high effort or the (nearly) impossibility to fake, delete or modify this type of traces.

In digitised forensics, based on the traditional forensic sciences and in biometrics, the evidentiary value is expressed in the light of two mutually exclusive hypotheses [MeVe12, p. 211] about a piece of data in digitised forensics mostly used for individualisation purposes (see also Section 2.2.2). But also the association process can stop at the classification step, e.g. for microtrace evidence in the shape of hair, where a reliable individualisation is typically not achievable. In criminal cases the first hypothesis typically represents the theory of the prosecution, namely that the digitised data has its origin in a trait of the accused (e.g. fingerprints). The second hypothesis, however, states the exact opposite (typically represented by the defence). The evidentiary value of the evidence of that digitised data is expressed as a ratio of probabilities of each of those hypotheses to be true in the context of relevant background information of the case [MeVe12, pp. 211-212], also known as the *likelihood ratio (LR)*. Together with a prior probability ratio the likelihood ratio will determine the posterior likelihood ratio. LR-based systems can provide statistical probabilities on the set of distinctive and automatically extracted features.

A very important precondition is that the background information is correct and that the methods used to calculate the respective probabilities. Also, the quality of the inference is highly depend-

ent on the quantity and the properties of the data used to estimate the intra source and inter source variability as with any biometric system (see [Vie06, p. 22]). Further, the enforcement of the security aspects, especially authenticity and integrity, of all data used throughout the biometric pipeline (see Section 2.8) greatly influences the evidentiary value of the digitised data and the inferences drawn from it.

2.8 The generic biometric pipeline for biometric user authentication

IT systems are increasingly used to assist the examinations of traditional forensic sciences (digitised forensics). A number of techniques used in traditional forensics aim at identifying or verifying the presence of people at the scene of crime according to the traces they left (e.g. fingerprint traces, microtraces, DNA). The discipline of *biometrics* also aims to identify or verify human beings based on distinctive physical or behavioural traits but uses automated methods in order to achieve those goals [PaMi10, p. 1]. However, the results are typically attached with a particular error rate, which renders this automated process insufficient to meet the high demands placed on the data, the processing and the result in a forensic setting. However, as pointed out in [MeVe10, pp. 207-208], both disciplines can learn a lot from each other.

In an attempt to digitise forensic proceedings, existing approaches from biometric systems (e.g. used for user authentication) are researched (see e.g. [JJNN08, pp. 1-8] or [HKG+11, pp. 1-8]). Such biometric systems typically employ the techniques of *pattern recognition* [Vie06, pp. 29-30]. Pattern recognition as stated in [DHS01, p. 1] is the act of taking in raw data and taking an action based on the "category" of the pattern. Pattern recognition approaches, according to [Bebi12, p. 11], can be categorised into:

- *Template matching* - The pattern to be recognized is matched against a stored template while taking into account all allowable pose (translation and rotation) and scale changes,
- *Statistical pattern recognition* - Focuses on the statistical properties of the patterns (i.e., probability densities),
- *Structural Pattern Recognition* - Describe complicated objects in terms of simple primitives and structural relationships,
- *Syntactic pattern recognition* - Decisions consist of logical rules or grammars. Artificial Neural Networks Inspired by biological neural network models.

A detailed description of the principles of those approaches would be outside the scope of this thesis. However, further details about the named approaches can be retrieved e.g. from [Bebi12, pp. 12-21]. The information flow during pattern recognition can be represented by a pipeline (Figure 8). Note that in literature, the terms pattern recognition and pattern classification are often used interchangeably and in this thesis both are used describing the same process.

Figure 8: Pattern classification pipeline (modified from [DHS01, p. 2])

During the *pre-processing* stage, operations are executed on data digitised from a sensor in order to simplify subsequent operations without losing relevant information (see [DHS01, p. 2]). Such operations can, for example, include image enhancement techniques such as noise filtering, contrast manipulations etc. One very important part of pre-processing is the *segmentation* operation. During this operation, different objects are separated from one another and from the background [DHS01, p. 2].

During the *feature extraction* stage operations are used to characterise an object to be recognised by measurements, whose values are very similar for objects in the same category and in the contrary, very different for objects from a different category [DHS01, p. 7]. In other words, the idea

is to look for distinguishing features (i.e. properties) with a high discriminatory power. A very important demand placed onto those distinguishing features is that of having a high degree of *invariance*. That means, those features have to be robust against common operations such as scaling, rotation, translation, occlusion, projective distortion, just to name a few (see [DHS01, p.11]). Typically a large number of features are represented as a *feature vector*, describing the measured values of the particular features of a given object. Those feature vectors are typically of a high dimensionality. Therefore, often this stage includes the process of *feature selection*, in which the most valuable features are chosen from a number of candidate features, resulting in a smaller feature vector, in order to reduce the complexity of the classification.

During the *classification* stage, operations are used to assign a feature vector and thus an object to a category [DHS01, p. 12]. An important problem is the relationship between feature values from objects belonging to the same category (intra-class variance) and the feature values from objects from different categories (inter-class variance). Obviously, for a pattern classification system the goal is to minimise the intra-class variance and to maximise the inter-class variance. One important cause of intra-class variance in real-world systems is *noise*. As defined in [DHS01, p. 12], noise in this context is a property of the sensed property, which is not due to the true underlying model. Instead, it is the result of the randomness in the world or the digitalisation process inside the sensor. It is thus closely related to the *uncertainty* as introduced in Section 4.1 of this thesis. In some pattern classification systems, also a post processing is performed. Post processing can include the addition of context, as defined in [DHS01, p. 13] as input dependent information from sources other than from the target pattern itself. In some systems, there is a feedback loop and the system can step back to stages already passed, which is depicted by leftward arrows in Figure 8.

At the heart of pattern recognition is an underlying set of *models*. Those models represent, according to [DHS01, p. 2] different descriptions (typically in a mathematical form). The main goal of pattern recognition is therefore to hypothesize the class of the models and to choose the model that corresponds best to the given pattern recognition task. The process of choosing the right model is also called training. In it, a training set of data is fed into the pattern recognition pipeline and evaluated. The classification results are used to alter the models until a given level of classification performance is achieved. This step of creating and evaluating models is typically executed ahead of a set of classification tasks, which is further discussed in Section 4.4 of this thesis. As stated before, pattern recognition is a highly complex problem and in this thesis only a brief description is provided. For further details the reader is pointed to [DHS01] and [Bebi12] as starting points for much more detailed information.

As stated in the beginning of this section, pattern recognition is one important part in the larger process employed in *biometric systems*, which are used in a number of biometric domains (e.g. for user authentication, convenience biometrics, medical biometrics, forensic biometrics, see [Vie06, pp. 13-15]). These systems exploit unique traits of humans that allow identifying or verifying a particular individual. Those traits form several modalities, which can be grouped into active and passive modalities [Vie06, p. 36]. Examples for active modalities (behavioural traits) include voice, handwriting, keystroke behaviour, which require the active cooperation of the individual. Of particular interest in the context of digitised forensics are the passive modalities (physiological measurements) such as Iris, Retina, Ear and Fingerprint since they typically do not ask for active cooperation of the individual. For a biometric system to work, two fundamental operational modes are needed, *enrolment* and *authentication* (see [Vie06, pp. 19-21]). In the operational mode of *enrolment*, the users are registered within the system. In the operational mode of *authentication*, a comparison between the stored data and the data acquired live from the person is executed. In this context, the authentication can be further divided into *verification* and *identification*. In a *verification* scenario, a binary decision is taken by the system, whether the presented data matches a declared identity. During *identification*, the presented data is matched against the set of stored enrolment data resulting in the selection of an individual out of the set of enrolled (registered) individuals.

The processing inside a biometric system can be depicted as a *biometric pipeline*, which is shown in Figure 9.

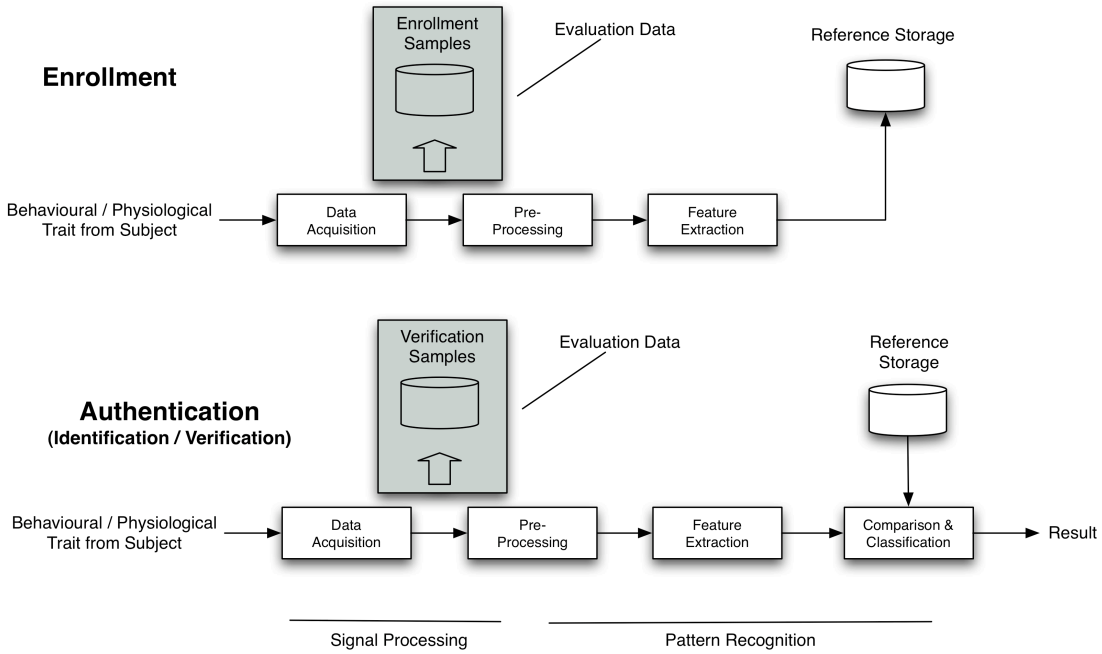

Figure 9: Biometric Pipeline in the operational mode of enrolment and authentication (modified from [Vie06, pp. 19-20])

In the stage of *data acquisition*, comprises according to [Vie06, p. 19] of the measurement and analogue-digital conversion of behavioural or physiological traits. For example in the case of the modality of fingerprint, optical and non-optical sensors (such as capacitive or pressure sensitive piezo-electric sensors) are used to create a two dimensional array, which comprises of scalar measurement results [Vie06, p.55]. Signal processing is used during data acquisition to shape the digitised representation of the analogue input.

In the stage of *pre-processing*, the acquired digital data is enhanced to support the following feature extraction using signal processing. In the exemplary chosen modality of fingerprint, such operations can include the histogram equalisation, direction-based filtering (e.g. Gabor filter) but also standard image processing operations such as smoothing, sharpening or binarisation [Vie06, pp. 55-56].

In the stage of *feature extraction*, the pre-processed data is examined for the presence of features, which allow for the comparison and classification in the following stage. In the exemplary chosen modality of fingerprint, this could include the minutiae detection. This is a feature present in every fingerprint and is a result of the friction ridge structure (see [Mac11a, pp. 2-3 - 2-26] for more details). Of particular interest is the type of ridge line endings (e.g. termination, bifurcation etc.) As stated in [Vie06, p. 56], on approach to feature extraction could be the definition of 8-neighbour rules, whereby each pixel is analysed in the context of its eight neighbouring pixels.

In the stage of *comparison and classification*, features derived from the sample are compared to features stored in the reference storage from the enrolment process. The result of the comparison yields a classification, which can be a binary decision (verification) or a multi-class decision (identification). In the exemplary chosen modality of fingerprint, a minutia-based approach could comprise of the comparison of relative position of minutiae by computing a map of positions of each individual point to all other minutiae contained in a fingerprint image [Vie06, p. 57].

In some systems, for evaluation purposes, pre-processed data representing digitised signals acquired from the traits is stored as evaluation data (see Figure 9). Since these evaluation data contain the acquired digital signals, replay attacks could be staged in production systems. Further,

these data are highly person-related data and therefore are subject to data protection acts and the resulting regulations (see Section 2.7). Therefore, these data are only used for evaluation purposes and are not stored in production systems. In a simplified way, only taking the stages of the biometric pipeline into account, a biometric system is summarised by the following Figure 10.

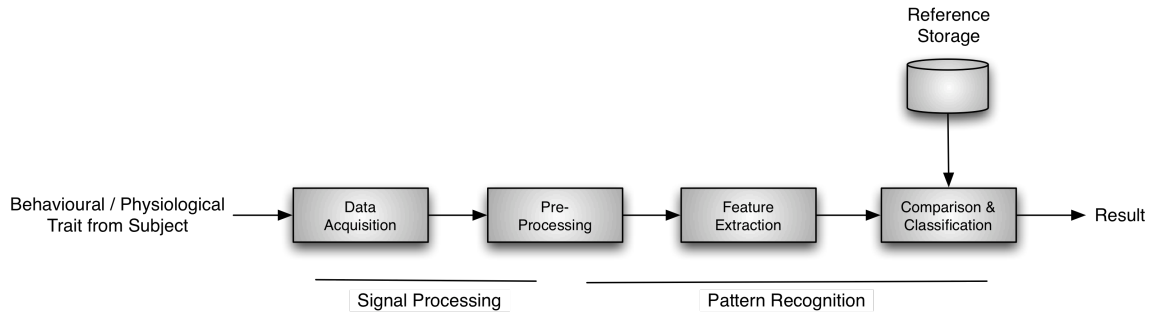

Figure 10: General model of a biometric system (modified from [Vie06, p. 30])

Notice how in [Vie06, pp. 29-30] the preprocessing is considered to be a part of signal processing, whereas in [DHS01, pp. 11-12] it is seen as part of pattern recognition. This can be consolidated by attributing the signal enhancement operations (e.g. noise filtering, contrast manipulation etc.) to signal processing and the segmentation operations of the pre-processing to pattern recognition. To derive techniques and methods for digitised forensics, it is vital to include the sensor and therefore the stage of data acquisition, as it is an important element of the chain of custody (see Section 2.4) for digital objects and for the chain of custody for physical objects (see Section 7).

Although some biometric systems (e.g. biometric systems using the modality of fingerprint or ear) share a lot of similarities with systems suitable for digitised forensics, there is a whole range of differences. To name only a few, there is the aspect of consent towards the acquisition of the traits and the cooperation of the individual. Both are highly unlikely to be the case of a perpetrator and the traces he leaves at the crime scene. The differences between biometric systems and approaches for digitised forensics are discussed in detail in Section 7 of this thesis.

2.9 Summary of necessary existing knowledge for the research challenge

In this Section 2 we started with the definition of our view of terms and definitions during which we arrived at a definition of a forensic examination that broadens the view by including the operators of IT systems as important sources of strategic preparation ahead of a specific incident, which can greatly enhance the amount and quality of data gathered, investigated and analysed as well as documented, should the IT system be met with an incident.

As part of our general fundamentals we looked at the properties of forensic examinations starting with traditional crime scene forensics on the basis of principles and processes. Their knowledge is important because the state of the art transferred it to digital forensics and back ported it to digitised forensics. It serves as a means to explain our chosen limitation and focusing on the data contained and communicated in IT systems as opposed to event reconstruction. In conjunction with setting objectives for forensic examinations (including definitely non-court based actions such as operational troubleshooting) it further motivates, as we will show later, how deep we have to dig into the data and which case-specific limits we set ourselves and it gives meaning to the separation of examination steps into different sets.

We motivated and described the necessity of maintaining a chain of custody and a comprehensive documentation. A dedicated examination step arises from that importance of those two subjects, which basically render a forensic examination into something worthwhile in the first place.

The questions a forensic examination aims at answering, based on a common knowledge to describe incidents in IT system, form the basis of describing a forensic examination as a set of distinct examination steps.

Anti-forensics serves as a man-made malicious source of loss, error and uncertainty in forensic examinations. We described the general goal and selected means to that end also to remind ourselves that forensic examinations often are conducted in an adversarial environment.

We looked at legal requirements from a technical and procedural perspective that render forensic examinations sensible (evidentiary value) and can limit the scope of such examinations.

We finished with the basis for the relatively new branch of digitised forensics (see [Sri10, p. 40]) and thus the use of IT systems to support crime scene forensics. One basis is the biometric pipeline describing the signal and data flow during enrolment and subsequent verification or identification and will use this pipeline as a basis to describe the process using a data-centric procedural view.

All general fundamentals are deemed necessary to understand the formalisation of loss, error and uncertainty in digital and digitised forensics and the data-centric process model to describe the examination itself and the occurrence of loss, error and uncertainty within.

We start by going into a detailed discussion about selected aspects of the state of the art in digital and digitised forensics and highlighting properties that are in support of our research challenge, namely (see Section 1.1): Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

3. Selected aspects of the state of the art in digital forensics and in digitised forensics

The contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Mario Hildebrandt, Claus Vielhauer, Robert Fischer, Christian Arndt (in descending order): [KHD10], [KHD+12], [FVH+13], [AKD+15]

In this section the state of the art in digital forensics and in digitised forensics is outlined. Digital forensics is used for about four decades [Pol10, p. 5]. It encompasses, according to [Wol09, p.3], approaches and techniques for gathering and analysing traces of human and computer-generated activity in such a way that it is suitable in a court of law. Its objective is to perform a structured examination into past and ongoing occurrences of data processing and transmission whilst maintaining a documented chain of custody for the evidence, which can be reproduced unambiguously and validated by competent third parties [Wol09, p.3]. However, the often practitioner-driven approaches are mostly supported by forensic tools from relatively small companies, which face extraordinarily high research and development costs. Because of rapid changing hardware- and software products, product lifetimes of such tailor-made commercial solutions are short [Gar10, p. 67]. This problem becomes very visible when looking at digital forensics for mobile phones, where a large number of hardware platforms and operating systems need to be covered by forensic tools. Even the mere data gathering is hindered by a multitude of connectors, inaccessible mass-storage (mostly flash memory directly surface soldered to the mainboard). Not describing the usage of a particular forensic software product, only a few best practice guides exist in that area (e.g. [JaA12, p6]), but there is no published standard way to extract information from mobile phones.
Compared to digital forensics, digitised forensics is still in its infancy and is currently the subject of active research. However, accepted procedures and methods used in traditional forensics (e.g. forensic dactyloscopy, see [Meu15, p. 729]) need to be adhered to and a mapping of those procedures and methods to the digital domain has to take place whilst exploiting the new possibilities (e.g. contact-less 3D fingerprint capture and analysis). All the legal requirements (see Section 2.7) need to be met by digitised forensics.

3.1 General existing forensic process models to cover the examination

In this section, existing models that describe aspects of forensic examinations are outlined. A large number of models used to formalise the forensic process exist. The following analysis describes important aspects of the respective process models together with unaddressed issues, which are picked up in the remainder of this thesis.

Although an exhaustive search for existing process models for digital and digitised forensics precluded and accompanied the writing of this thesis, it has to be stated, that the selection of presented models is of an exemplary nature and the models are chosen because of the fundamentals they introduce and of unaddressed issues they leave for additional research.

3.1.1 Cyber Forensic Assurance (CFA) [Dar10]

The model of the Cyber Forensic Assurance (CFA) by Glen S. Dardick [Dar10, pp. 61-64] describes from a practitioner's view the problem of four potential types of errors. Notice, that the notion of error in this context refers to a broader definition of errors typically made by the examiner, addressing mostly the process of reconstruction of events (see Section 2.2), as they include hypotheses of as to how events might have occurred. In Section 4.1 a formal definition of error is introduced, based on forensically relevant data in IT systems.

From Section 2.7, the “Null” hypothesis in forensic examination involving the legal system in a number of countries is typically that the defendant is innocent and the burden of proof lies with the prosecutor. Using this “Null” hypothesis, potential errors in forensic examinations, according to [Dar10, p. 58] can be classed into the following types of errors:

- Type I Error - False Positive: Incorrectly rejecting the "Null" hypothesis when in fact it is true is referred to as a Type I error, or a false positive. In effect, by rejecting the "Null hypothesis" it is falsely proven that the defendant is not innocent,
- Type II Error - False Negative: Not rejecting the "Null" hypothesis when in fact it is false is referred to as a Type II error, or a false negative. In effect, by not rejecting the "Null hypothesis" it is not proven that the defendant is not innocent.

Those types of error (false positive and false negative) are also relevant and well known in other fields of research such as pattern recognition and statistics (e.g. see [ThKo06, p. 734]). In addition to those types of errors, two other types of error are outlined in [Dar10, p. 63]:

- Type III Error - Having solved the wrong problem: Here it is argued that one of the most important determinants of a problem's solution is how that problem has been represented or formulated in the first place (originally in [MiFe74, pp. 383-393]),
- Type IV Error - Incorrect interpretation of a correctly rejected hypothesis: Here in an analogy it is stated that type IV errors can be likened to a physician's correct diagnosis of an ailment followed by the prescription of the wrong medicine (originally in [MaLe70, p. 398]).

In an attempt to qualify and minimise those aforementioned types of errors the Cyber Forensic Assurance Model (CFA) is introduced in [Dar10, pp. 61-64]. The security aspects of information security, information assurance as well as the Parkerian Hexad and information quality are assembled to the Cyber Forensic Assurance (Table 1):

CFA	Components
I	a) Confidentiality - ensuring that information is accessible only to those authorized to have access
	b) Possession/Control - i.e. chain of custody
II	a) Integrity/Consistency - perceived consistency of actions, values, methods, measures and principle - unchanged "is it true all of the time?" (Verification)
	b) Authenticity/Original - quality of being authentic or of established authority for truth and correctness - "best evidence" (Validity)
III	a) Availability/Timeliness - the degree to which the facts and analysis are available and relevant (valid and verifiable at a specific time)
	b) Utility/Relevance - "Is it useful/is it the right information?"
IV	a) Completeness - "Is it the whole truth?"
	b) Non-repudiation/Accuracy - transaction cannot be denied (Validity) - no alternate hypothesis

Table 1: Cyber Forensics Assurance model taken from [Dar10, p. 62]

The CFA contains the model of the *Parkerian Hexad* [Par97, pp. 14-19] and that of *information quality* [Mil96, pp. 79-82]. The term *Parkerian Hexad* is coined by [Kab12, pp. 4-17] in reference to its originator and adds the following items [Par97, p. 16]:

- Possession/Control of information - means to have information in hand or to control information under specified circumstances,
- Utility of Information - refers to the state of information being useful or fit for some purpose,
- Authenticity of Information - is the extrinsic state of being true, genuine, original, and valid by being in conformity to fact and reality.

According to [Mil96, pp. 79-80], *information quality* has ten dimensions, out of which in [Dar10, p. 60] five are chosen in the context of forensic examinations by asking relevant questions:

- Accuracy – Is the information factual?
- Relevance – Is it the right information?
- Consistency – Is it right all of the time?
- Timeliness – Is it applicable to the appropriate timeframe?
- Completeness – Is it the whole truth?

The Cyber Forensics Assurance model, as concluded in [Dar10, p. 64], is geared towards a better understanding about finding ways to conduct forensic examinations that compliant with a changing legal and technological landscape and is developed to assist in the training of Cyber Forensics professionals. It is believed to be a tool to reduce the error types outlined above.

Discussion

The model of the Computer Forensic Assurance CFA as presented in [Dar10, pp. 61-64] in the context of this thesis yields useful foundations, next to underlining the need for addressing the security aspects of confidentiality, integrity, availability, non-repudiation and authenticity, particularly by introducing the demand for possession/control, completeness as well utility. These demands introduced by the CFA, which are shown to highly relevant in forensic examinations, are used by the Data-Centric Examination Approach (DCEA) introduced in this thesis (see Section 4).

The main criticism towards the Computer Forensic Assurance CFA is that only goals are outlined (high level practitioner's view only) but no means to their enforcement are given.

3.1.2 NIST Special Publication 800-86 - Guide to Integrating Forensic Techniques into Incident Response [KCG+12]

The model introduced in [KCG+12, pp. 3-1 - 3-8] aims at adding forensic capabilities into incident response proceedings. The Special Publication 800-86 "Guide to Integrating Forensic Techniques into Incident Response" has the clear intention of providing an IT view instead of a law enforcement view [KCG+12, p. 1-1]. This is the equivalent of the view of the operator in the context of this thesis. Clearly to be classified as the high level practitioner's view according to [PBM08, pp. 116-117], the steps advised to be taken are grouped into four distinctive categories [KCG+12, pp. 3-2 - 3-7]. These categories are abbreviated and summarised for the intends and purposes of this thesis as follows:

- Collection - identification of potential sources of data, proactive collection of data (e.g. audits), implementation of centralised logging, data acquisition based on likely value, volatility, amount of effort required, acquisition by forensic duplication wherever possible, integrity verification using message digests and cryptographic hash sums, detailed documentation of each step, preparation of storage space and devices ahead of incidents, incident response considerations (e.g. securing perimeters of the affected IT system),
- Examination - involves bypassing or mitigating OS features (e.g. compression, encryption, access control), reduction and filtering of relevant data, inclusion or exclusion of files based on forensic databases,
- Analysis - study and analyse data to draw conclusions, identifying people, places, items and events, determination of relation between the identified data, correlation of data from multiple sources, integration of data from pro active measures (e.g. intrusion detection, security event monitoring),
- Reporting - preparation and presentation of information from the analysis, includes alternative explanations, identification of actionable information to allow collect new sources of information, audience consideration and tailoring of the report, identification and remedy of procedural shortcomings and policy errors, maintenance and growth of experts skills.

The following Figure 11 illustrates the four steps as a directed path of action with a feedback loop from the documentation to include improvements to the forensic process for all steps to be applied for the next forensic examination.

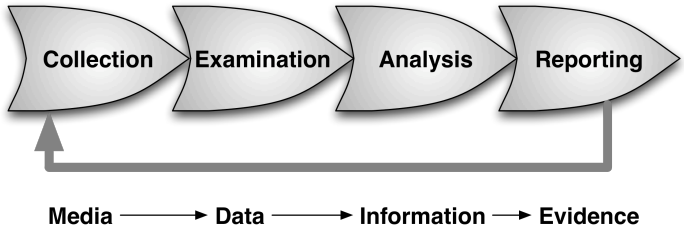

Figure 11: Forensic process categorised into four steps and the road from media to evidence (modified from [KCG+12], p. 3-1)

In addition to the grouping of examination steps, also the road from media over data to information and finally evidence is projected towards the respective steps.

Discussion

This high-level practitioner's view according to [PBM08, pp. 116-117] has its target audience mainly in the operators of IT systems [KCG+12, p. 1-1] in general and in particular for incident response teams, forensic analysts, system-, network- and security administrators as well as computer security program managers. In doing so, a very important property of IT operators, namely that they can administer measures ahead of a particular incident but in the anticipation of such, is not included in the modelling of the examination steps. This drawback is addressed in the Data-Centric Examination Approach (DCEA), which is at the core of this thesis, by including the examination step of Strategic Preparation *SP* (see Section 4.4.1.1).

Although somewhat similar to the transition from data to knowledge (see Section 2.1.2 of this thesis), the authors of this thesis refrain from mapping items such as data to a particular examination steps; simply because each of the named items according to the experience of the authors can appear in any of those steps. The model does not address the topic of loss, error and uncertainty in a structured, dedicated way.

3.1.3 The Four-Step Process according to [New07]

The process description of the Four-Step process in [New07, pp. 5-7] starts with the acquisition, thus omitting a dedicated preparation step. However, the description of the acquisition step includes many ideas that are part of the operational preparation, e.g. the weighing of the live-forensics vs. post mortem forensics (see also Section 3.3.1.2) or the initialisation of the chain of custody (see also Section 2.4). It also stresses the need for constant documentation throughout the whole process. The Four-Step Process separates the forensic process into distinctive steps, whose contents are abbreviated and summarised for the intends and purposes of this thesis as follows:

- Acquisition - differentiation between live and post mortem forensics, initialisation of chain of custody, data acquisition always with the preservation of any evidence in its original form in mind (i.e. avoid contamination of the data by the acquisition itself),
- Identification - evidence with both physical (hw/sw components, e.g. disk drive) and logical (address or location of evidence, e.g. on a disk drive) context, documentation of all information relating to the steps taken,
- Evaluation - examiner to evaluate the relevance and validity of the collected evidence, making sure that a valid chain of custody exists and integrity has been kept, separation of case-specific and irrelevant data,
- Presentation - decision by the forensic team as to the worthiness of various pieces of evidence, presentation of both sides of an argument.

The model allows for some steps to be repeated in the light of new findings (see Figure 12).

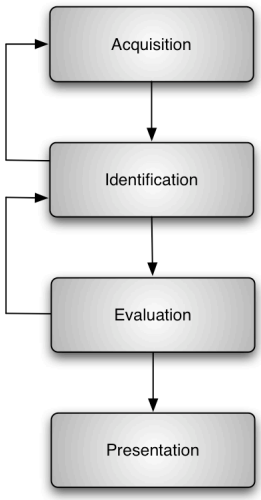

Figure 12: Four step model of the forensic process (modified from [New07, p. 6])

Stepping backwards from evaluation all the way back to the acquisition is possible. This is included for examinations during which new facts come to light during the evaluation that point to other data sources (e.g. removable mass storage devices etc.) or to new findings that need to be re-examined during the identification step.

Discussion

The analysis of the Four-Step Process model of [New07, pp. 5-7] in the context of this thesis yields insights into the potentially re-visiting of examination steps (see Section 4.4.1.7). The model description in [New07, p. 6] mixes physical and digital evidence. But it doing so it inspires the need for separation of data or the Data-Centric Examination Approach (DCEA), i.e. which data is part of the digital data contained in the system and which is added by the examiner from external sources (e.g. printed serial numbers, model numbers etc., see Section 4.4.2.2). Generally, it also acts as a tool to check for the completeness of the Data-Centric Examination Approach (DCEA) with regards to traditional view of digital forensics. This high-level practitioner's view according to [PBM08, pp. 116-117] has its target audience mainly in the incident responder and crime investigation community.

This model lacks the computer scientist's view on digital forensics according to [PBM08, pp. 116-117]. Further, it does not contain a dedicated discussion, let alone structuring, of loss, error and uncertainty.

3.1.4 The basic methodology according to [KH02]

This model describes a forensic examination by grouping it into three high level steps [KH02, pp. 5-20]. Although a dedicated preparation step is not included in the model itself, a number of aspects are mentioned in the description of the acquisition step. The model is part of a larger approach to offer hands-on advice when the necessity of an examination using computer forensics is required. Its main intention is to introduce its users to the general concept of computer forensics and draws parallels to the examination of items in traditional crime scene investigation. This connection, however, is not as detailed and formalised as in the model from [CaS03, pp. 5-13] to be discussed in Section 3.1.7 of this thesis. The model introduced in [KH02, pp. 5-20] separates the forensic process into three distinctive steps, whose contents are abbreviated and summarised for the intends and purposes of this thesis as follows:

- Acquire the evidence - initialisation and maintenance of a chain of custody, collection of evidence, identification of evidence with entry and signature on the chain of custody, photographing of the scene, transportation of evidence with sealing of the package, storage of sealed evidence in secure area with limited access, documenting of the steps taken,
- Authenticate the evidence - cryptographic hash sums and timestamping of electronic evidence,
- Analysis - view of filesystem meta data, hex editor view of suspect files, unerase of deleted files, keyword search, unallocated and slack space, preservation of evidence, court presentation

As part of the acquisition step the concept of the chain of custody is outlined (see also Section 2.4). The chain of custody is basically a mechanism to answer the following questions:

- Who collected it?,
- How and where?,
- Who took possession of it?,
- How was it stored and protected in storage? and
- Who took it out of storage and why?

It has to be noted that in the context of the model from [KH02, pp. 5-20], this only applies to physical, tangible items (e.g. IT devices, mass storage devices etc.).

Discussion

The model implements the practitioner's view according to [PBM08, pp. 116-117] and has its target audience mainly in newcomers to the field of digital forensics. The general concept of the chain of custody is included into the examination step of operational preparation and further extended for use in the Data-Centric Examination Approach (DCEA) introduced in this thesis. The acquisition is preceded by a weighing process concerning the operational preparation of a forensic investigation with regards to keeping the system running during a suspected incident or to shut it down. Further, if the conclusion is geared towards shutdown, the means of shutdown (orderly vs. power cord/main battery disconnect) needs to be weighted and the conclusion properly justified and documented. This is included in the operational preparation step in Section 4.4.1.2 for the Data-Centric Examination Approach (DCEA) introduced in this thesis. More generally, it also acts as a tool to check for the completeness of the Data-Centric Examination Approach (DCEA) with regards to traditional view of digital forensics.

This model, by its very nature and intention, lacks the computer scientist's view on digital forensics according to [PBM08, pp. 116-117] and a dedicated discussion, let alone structuring of loss, error and uncertainty.

3.1.5 Applying Traditional Forensic Taxonomy to Digital Forensics [Pol08]

This model takes in the ideas of *principles* and *processes* discussed by [InR01, pp. 77-78] when addressing traditional forensic examinations (see Section 2.2). The author arrives at the conclusion that the principles (such as Locard's exchange principle, see Section 2.2.1) are not applicable in the field of digital forensics. However, the processes as described in [InR01, pp. 77-78] are used to form a model of the forensic process for digital forensics. The mapping results in four distinct examination steps is abbreviated and summarised from [Pol08, pp. 20-23] for the intends and purposes of this thesis as follows:

- Identification - helps to describe digital evidence in terms of its context (*physically* e.g. particular brand of harddrive, *structurally* e.g. number of cylinders, heads, sectors, *logically* e.g. FAT32 Partition, *location* e.g. directory or file, *content* e.g. memo, spreadsheet, email),
- Classification/Individualisation - *characterisation into classes* e.g. file systems, partitions, individual files -> common origin, *individualisation* by mathematical signature e.g. cryptographic hash value,

- Association - in digital forensics necessity of identification of items (e.g. files, data, structures, code) that need association, determination, where such items are likely to be found and what tools to be used for their location,
- Reconstruction - "ordering of associations in space and time" [InR01, p. 79], answers to the question of where, how and when (mostly relative time only).

In [Pol08, p. 23] the author further discusses the hard problem of formulating scientific questions only after the legal/investigative questions are established. He states that often forensic examinations are merely based on what could be technically retrieved and analysed (i.e. the scientific questions) and this basically sets the stage for the investigative questions to be formulated. The author demands for a reversal and thus for an information-centred approach, which is integrated into the model proposed in [Pol08, pp. 20-23].

Discussion:

The analysis of the model from [Pol08, pp. 19-25] in the context of this thesis yields one the basic principle of data-centric thinking as used for the Data-Centric Examination Approach (DCEA) introduced in this thesis. As outlined in Section of this thesis, the source for information is data. By placing the focus onto data on the system under examination as opposed to application-specific capabilities of software, the Data-Centric Examination Approach (DCEA) represents a step towards the demand for information centred approaches. This demand is used to back the introduction of the forensic data types in Section 4.2.

The model represents a high-level practitioner's view according to [PBM08, pp. 116-117] and describes a forensic process divided into examination steps. It is addressing the standpoint of law-enforcement and thus omits the strategic preparation. The model does not address the topic of loss, error and uncertainty.

3.1.6 A Common Process Model for Incident Response and Computer Forensics [FrS07]

The general idea of this model is to incorporate forensic principles and procedures into incident response during the analysis phase [FrS07, pp. 19-40]. It describes a holistic approach that aims to unite incident response and computer forensic processes. In reference to [PBM08, pp. 116-117] it provides a practitioner's view for the forensic process since it is introducing four broad examination steps together with extra detail. The steps are abbreviated and summarised for the intends and purposes of this thesis as follows (see also Figure 13):

- Pre-Incident Preparation - ongoing phase, takes place *before* an incident, preparation of the staff (organisation or CSIRT), implementation of host- and network-based security measures, organisational measures
- Pre-Analysis Phase
    - Incident Detection - occurs whenever a person or security mechanism suspects an unauthorised or unlawful action involving a computer system or network, emphasis on comprehensive documentation
    - Initial Response - determination of type and scope, containment measures to limit the potential damage of an ongoing incident
    - Formulation of Response Strategy - determination of the most appropriate strategy, decision about a fully fledged forensic examination, judging the criticality of compromised hosts/data and potential perpetrators and apparent skill level of the attacker and downtime and monetary loss

- Analysis Phase
    - Live Response - collection of mostly volatile data (see Section 3.3.1.1) whilst modifying the IT system as little as possible
    - Forensic Duplication - exact copies of all storage media involved in the incident without altering the original evidence, initialisation of a chain of custody for all media, safe storage of the original media and duplicates
    - Data Recovery - restoration of deleted, damaged, hidden or otherwise inaccessible data
    - Harvesting - gathering of metadata about the preserved material, structuring of largely unorganised material based on timestamps, permissions or other attributes
    - Reduction and Organisation - elimination of data deemed case-irrelevant, organisation of remaining structured data, organisation to allow efficient searches, identification and reference to relevant data (Analysis and Post Analysis)
    - Analysis - detailed reconstruction of the event comprising the incident, correlation of pieces of evidence to establish links, testing of multiple theories together with attempts to disproving them (elimination), analysis results are repeatable, comprehensive documentation
- Post-Analysis Phase
    - Report - describes details of the incident, is understandable to non-technicians or executives and meets the legal standards for being admissible in court, takes in all documentation for a comprehensive overview
    - Resolution - containing, solving the incident and prohibiting the recurrence in future

The interconnections of the phases within the whole of the model are shown in Figure 13.

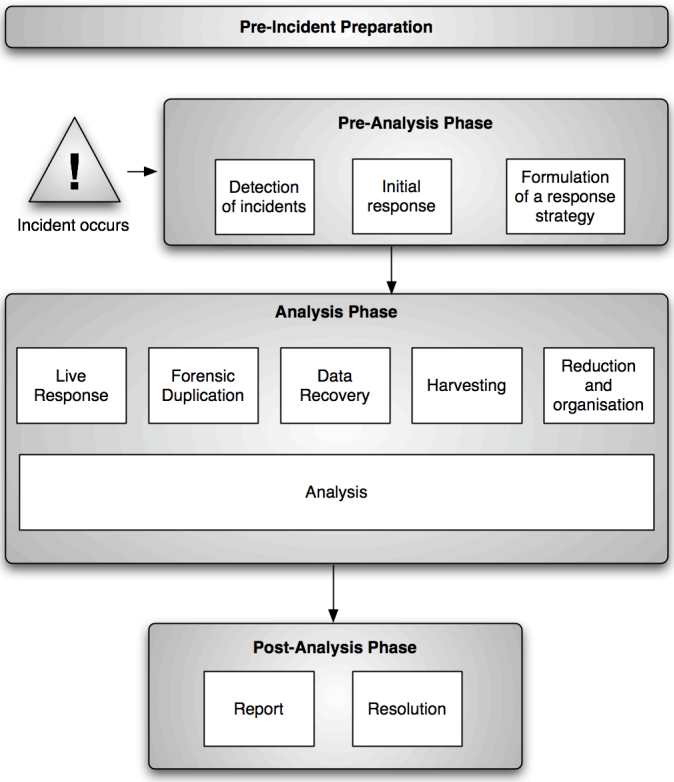

Figure 13: Common Process Model for Incident Response and Computer Forensics (modified from [FrS07, p. 29])

This model introduces a live response to the proceedings of an examination using digital forensics (see also Section 3.3.1.2). From the depiction, backward steps are not included. One further interesting detail of [FrS07, p. 39] is the determination of the soft factors of attacker thread level and potential damage to decide whether a full-scale forensic examination should be conducted. Those case-specific soft factors are proposed to be used to determine a threshold for the execution of the full-scale examination.

Discussion

This analysis of this model in the context of this thesis yields primarily support for the argumentation for the introduction of the case-independent examination step of strategic preparation (see Section 4.4.1.1) to the DCEA model introduced in this thesis. In doing so, a very important property of IT operators, namely that they can administer measures ahead of a particular incident but in the anticipation of such, is addressed in the modelling of the examination steps. It also backs the necessity for the actions used for the remaining examination steps of the Data-Centric Examination Approach (DCEA, see Section 4.4.1). Further, we expand on the thought of integrating a live response into the forensic process. This idea is picked up when structuring the sets of examination steps when live response, read live-forensics (see also Section 3.3.1.2), is spread across the examination steps (see Section 4.4.1).

Since the model from [FrS07, pp. 19-40] is not concerned with the computer scientist's view according to [PBM08, pp. 116-117], it makes no statements regarding the structuring of data contained in the IT system under examination nor to the structuring of the structuring of the methods to identify, gather, investigate, analyse or document the traces contained in the data. The model does not address the topic of loss, error and uncertainty.

3.1.7 Getting Physical with the Digital Investigation Process [CaS03]

In [CaS03, pp. 1-20] the authors married the physical and the digital investigation process to form one holistic model. The general idea is to treat the system under examination as a crime scene and thus requiring crime scene investigation methods. This view differs from the traditional view mostly by acknowledging the likeliness of vast amounts of potential evidence that can be examined by very different methods from which very diverse information can be gained. For example, each piece of digital evidence can be analysed and potentially yield information towards identifying ownership, location, and timing.

Ahead of their structuring of the holistic model from the viewpoint of crime scene investigation they place 5 very important demands on the model itself [CaS03, p.1]:

- must be based on existing theory for physical crime examinations,
- must be practical and follow the same steps an actual examination would take,
- must be general with respect to technology and not be constrained to current products and procedures,
- must be specific enough that general technology requirements for each phase can be developed,
- must be abstract and apply to law enforcement examinations, corporate examinations, and incident response.

Very notable are the demand for generality with regards to the technology used and the demand for being specific enough to be able to develop requirements for each phase. The model from [CaS03, pp. 7-12] suggests five groups of phases containing 17 phases in total (see Figure 14 also for a meta view).

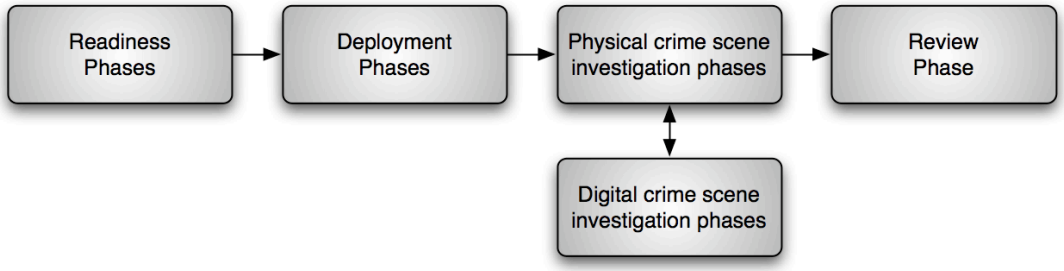

Figure 14: The five groups of phases in the investigation process (modified from [CaS03, p.7])

Those groups of phases together with the constituent phases are abbreviated and summarised for the intends and purposes of this thesis as follows:

- Readiness Phases - ensure that operations (training and equipment for personnel) and infrastructure (e.g. centralised secured logging, NTP time synchronisation, hash database) are able to fully support an examination, ongoing phase and not specific to a particular incident, separation into Operations Readiness Phase and Infrastructure Readiness Phases
- Deployment Phases - provide mechanisms for incident detection (manual and automated approaches) and confirmation (authorisation to proceed with further examination)
- Physical Crime Scene Investigation phases (in parallel with Digital Crime Scene Examination)
    - Preservation Phase - secures the crime scene (identical for digital and non-digital crimes), can include identifying people that are present and denying access to the scene to others
    - Survey Phase - identification of pieces of physical evidence (obvious and fragile), development of an initial theory about the crime, immediate documentation and collection of fragile pieces of evidence
    - Documentation Phase - documenting the crime scene and physical evidence using sketches, photographs and videos, capture information about the layout and important details (incl. connections to IT systems and state information, serial numbers)
    - Search and Collection Phase - in-depth search of the scene for additional physical evidence and its collection, including methodical (strict search patterns) searches
    - Reconstruction Phase - organisation of analysis results from the collected physical and digital evidence, correlation of physical and digital evidence to link a person to an incident
    - Presentation Phase - presentation of physical and digital evidence to court or corporate management, presents evidence and theory from the crime scene reconstruction
- Digital Crime Scene Examination phases (each digital device is a separate crime scene)
    - Preservation Phase - securing the entrances and exits to the digital scene, i.e. the entire digital environment (e.g. isolation from the network, collection of volatile data, identification of suspicious processes), complete forensic images of IT systems to be analysed by the lab later
    - Survey Phase - identification of obvious pieces of digital evidence (case-specific), identifies the skill level of the suspect and what analysis techniques are required
    - Documentation Phase - proper documentation of all digital evidence found, employment of cryptographic hash sums to prove Integrity

- Search and Collection Phase - thorough analysis of the IT system for digital evidence (e.g. keyword search, extraction of data from unallocated space, restoration of deleted files)
- Reconstruction Phase - involves putting together the pieces of the digital puzzle, classification and assessment of digital evidence to determine the amount of trust that can be placed in it, employment of advanced analysis techniques (e.g. executable analysis, decryption)
- Presentation Phase - presentation of the digital evidence found on the physical scene, integration of the results of each digital crime scene, often and ongoing share of information between the physical and the digital team

- Review Phase - identification of areas of improvement, results are new procedures, new training

The interfacing of the digital crime scene investigation with the physical crime scene investigation is depicted in Figure 15.

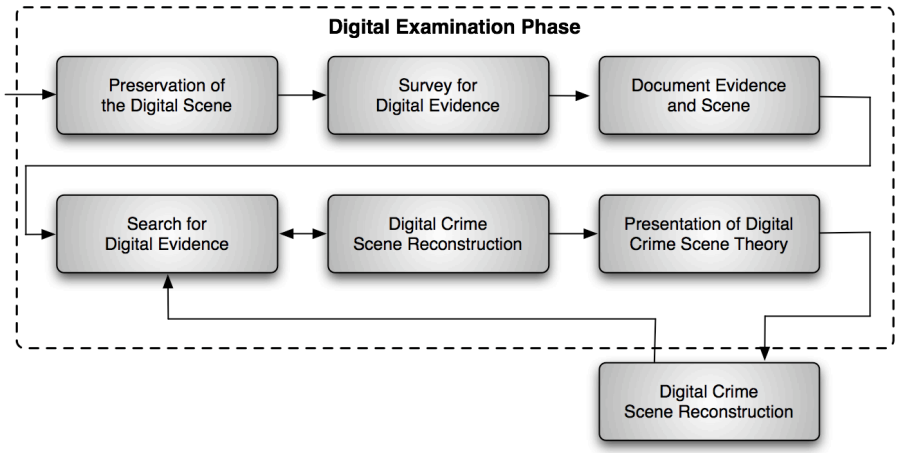

Figure 15: The six phases in the digital crime scene investigation. The results are fed back to the physical crime scene investigation (modified from [CaS03, p. 10])

Of particular relevance for the context of the Data-Centric Examination Approach (DCEA) introduced with this thesis is the digital crime scene investigation phase. It reflects the crime scene approach by, amongst others, stating that each digital device is a separate crime scene. Interesting is the documentation phase which only addresses the documentation of the evidence itself (i.e. source and result) but not explicitly the means by which the result was derived from the source. Interesting is that only one back step from digital crime scene investigation is allowed. If during the search or the reconstruction signs for new digital elements of the digital crime scene surface, they are not covered in this model.

Discussion

The analysis of the model from [CaS03, pp. 5-13] yields a number of important findings for use in the Data-Centric Examination Approach (DCEA) introduced in this thesis. First there is the detailed description of the readiness phase and the separation of the Operations Readiness Phase and the Infrastructure Readiness Phase therein, both which is included in the examination step of strategic preparation in Section 4.4.1.1. Further, the demand for generality and independence from current product and procedures is one of the motivations for devising the model introduced in this thesis and shapes the research question introduced in Section 1.1. Instrumental for the model is the idea of a digital crime scene consisting of a number of digital devices themselves consisting of a number of components, each of which contain data that can be examined under a number of aspects (e.g. ownership, location, timing).

The model represents a high-level practitioner's view according to [PBM08, pp. 116-117] and describes a forensic process divided into examination steps. The model does not intend to address the topic of loss, error and uncertainty. Clearly, crime scene investigations and the associated methods and procedures play a very important role in the Data-Centric Examination Approach (DCEA). However, we want to broaden the general view and want to devise a model that is equally suited for applications where no criminal or malicious intentions are involved (e.g. hardware- and software faults, configuration mistakes etc.).

3.1.8 Computer Forensics Field Triage Process Model (CFFTPM) [RGM+06]

The Computer Forensics Field Triage Process Model (CFFTPM) according to [RGM+06, pp. 27-40] is a practitioner-driven approach for incidents, where time is at the essence. Whilst retaining demands placed onto forensic examinations to keep the integrity of the data collected, the authors make deliberated curtailments regarding the completeness (see also Section 3.1.1) of the collected data. The suggested model is never intended to fit all situations; it justifies the curtailment for incident where time is crucial (e.g. leads to follow abduction cases). Similar to live-forensics (see Section 3.3.1.2), the execution of triage examinations on site does not rule out a following full examination in the lab afterwards. However, the triage is likely to have an effect on the data stored in the system under examination as it contains 5 out of 6 steps *on site* (as opposed to a forensic lab):

- Planning [RGM+06, pp. 31-32] - the examiner has a matrix that *quantifies* the various possibilities on site (crime scene, suspect, digital evidence), uses pre-emptive case intelligence, the authors suggest a mapping of matrix to Situation Paragraph of a military Operations Order (OpOrd) with *SALUTE:*
    - *S*trength (Suspect count/suspect capabilities),
    - *A*ctivity (specific actions of the suspect),
    - *L*ocation (physical and virtual location),
    - *U*niform (deterministic symbols, e.g. email address, URL, network domains),
    - *T*ime (chronological scope for investigative searches based on previously gathered case intelligence),
    - *E*quipment (expected hardware devices incl. networks and software applications).

  The lead examiner has to make many specific decisions *prior* to arriving at the crime scene, suited hard- and software tools for *field examinations* are an absolute prerequisite.
- Triage [RGM+06, pp. 32-33] - A process where things are ranked in terms of importance or priority, i.e. pieces of or containers of evidence according to importance or order of volatility including the identification and prioritisation of potential containers of evidence, reaching a decision based on potential relevant evidence that can be obtained in a reasonably short time frame or data with a high volatility and involving a direct cooperation between examiners and interviewers.
- Usage/User Profiles [RGM+06, pp. 33-34] - Here the examination and analysis are executed, which includes placing artefacts in context with real-world events in the same manner as in "traditional IT forensics" but in field triage a very fast determination if the examination/analysis can be done within time constraints, which requires justified assumptions and thorough knowledge of *user profiles* and *artefacts relating to usage* (e.g. on Microsoft Windows based PC, where it consists of files, folders, registry keys and file properties.

- Chronology/Timeline [RGM+06, pp. 34-35] - digital evidence is defined by its temporal value (i.e. MAC times) and MAC times represent a *possibility to qualify the searches* and thus *quantify the evidence*:
    - 1st quantification - identification of time periods of normal use of the IT system,
    - 2nd quantification - applications and files used during *qualified times of interest*,
    - 3rd quantification - identification and analysis of recent shortcuts and stored information.

  However, the problem of drifting clocks and manipulated timestamps persists and needs to be taken into account.
- Internet [RGM+06, pp. 35-37] - almost every criminal action is somehow also reflects on internet activity, therefore browser (e.g. cookies, history), E-Mail and Instant Messaging artefacts should be examined and analysed, esp. E-Mail and Instant Messaging analysis might take enormous time therefore careful weighing of expected benefit vs. needed time
- Case-specific evidence [RGM+06, pp. 37] - forensic examiner should evaluate time resources, utilise pre-raid intelligence, customise search goals and prioritise search goals, differentiation between *bounded* (definite deadline for halting the search or loss of evidence value) and *unbounded* (as soon as possible but the quicker the better) time constraints, time costs of any examination activity vs. potential for fruitful results, it is generally best to perform tasks which can be accomplished quickly first

The CFFTPM approach is designed to be a real-world application of digital forensic examinations from the viewpoint of law enforcement agencies. It includes both live forensics and the more traditional post-mortem forensic approaches (see Section 3.3.1.2), both of which are to be executed on-site using integrity preserving measures (e.g. presence of forensic workstation, write-blocker for mass-storage devices etc.).

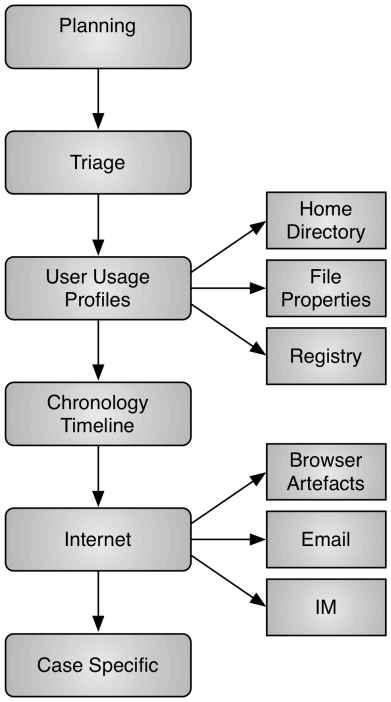

Figure 16: Examination steps for the Computer Forensics Field Triage Process Model (CFFTPM), modified from [RGM+06, p. 29]

The steps are (according to [RGM+06]) devised to be run through in a linear fashion (see Figure 16) with no dedicated mention of backsteps in the light of the results of each step. This is likely to be the result of time pressure that indicated the use of the CFFTPM approach in the first place.

Discussion

The analysis of the model of the Computer Forensics Field Triage Process Model (CFFTPM) in the context of this thesis yields more insight in the criticality of the step of operational preparation *OP* (see Section 4.4.1.2) of an examination in digital forensics. Interestingly the problems associated with time (absolute and relative) are mentioned, but no concrete measures are suggested. We discuss this problem in Section 3.3.1.4 and address it in all of the examination steps of the Data-Centred Examination Approach (DCEA) in Section 4.4.1.

The model represents a high-level practitioner's view according to [PBM08, pp. 116-117] and describes a forensic process divided into examination steps. The model does not intend to address the topic of loss, error and uncertainty. It uses the police raid-style approach and does not address the view of the operator of IT systems and thus does not mention strategic preparation. However, it describes the less-than-ideal circumstances under which a number of digital forensics examinations are executed.

3.1.9 A hierarchical, objectives-based framework for the digital investigation process [BeC05]

The model (framework) presented in [BeC05, pp. 148-153] can be seen as a meta-model in that it incorporates the results of a comparative study of existing forensic process models. It re-arranges existing process steps and adds binding principles that are covering the whole process and determine the steps to be taken during the examination (see Figure 17).

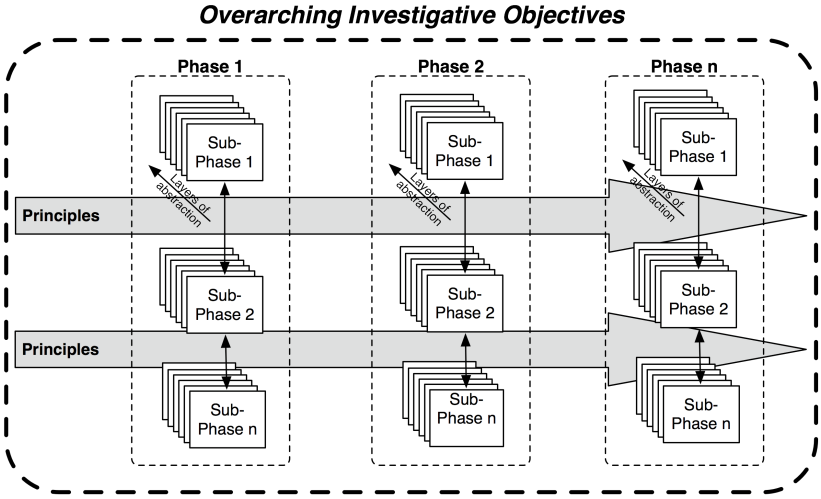

Figure 17: Hierarchical, objectives-based framework for the digital investigations process (modified from [BeC05, p. 149])

Generally, according to [BeC05, p. 149], the main constituents are:

- Phases (First Tier),
- Sub-Phases (Second Tier),
- Principles,
- Objectives.

Phases as first tiers (see [BeC05, p. 149] describe distinct, discrete steps, usually a function of time, suggesting a necessarily sequential and sometimes iterative approach. The phases are largely supposed to be non-iterative. They result from a comparative study of examination steps described in other forensic process models. The authors in [Bec05, pp. 149-151] follow the two paths of incident response and forensics and arrive at the distinct phases of:

- Preparation - focused on the organisation targeted by the investigation, known as forensic readiness, entails measures to improve the quality and availability of digital evidence whilst minimising organisational cost and burden,
- Incident Response - detecting, validating, assessing and determining a response strategy for the suspected security incident,
- Data collection - to collect digital evidence in support of the response strategy and investigative plan, it includes live-response, data collection from hosts and networks whilst ensuring integrity and authenticity of any data gathered,
- Data analysis - most complex and time-consuming phase, transformation, extraction and reduction of data collected, confirmatory analysis, event reconstruction based on the evidence gathered,
- Findings presentation - communication of relevant findings to various audiences to provide succinct and detailed confirmatory and event reconstruction,
- Incident closure - to preserve knowledge gained to enhance subsequent investigations, includes critical review of the entire process, implement enhancements of the process, dispose of evidence (return to owner, destruction cleaning and re-use), preservation of all incident-related information.

In a finer granularity, *sub-phases as second tiers* (see [BeC05, p. 153], describe distinct, discrete steps, usually a function of time, suggesting a necessarily sequential and sometimes iterative approach. They share these characteristics with first tiers but on a smaller scale. Sub-phases are supposed to include all types of crime and digital evidence and consist of tasks and task hierarchies subordinate to specific objectives of interest.

Overarching the whole examination are *digital investigation principles* (see [BeC05, p. 152]. They define procedures, guidelines and/or methodological approaches overlapping all phases and sub-phases. By definition they are not distinct, discrete phases or steps. They can naturally translate into constraints. Principles represent goals (e.g. proper documentation) One important principle is the evidence presentation principle. Its goals are the maximisation of evidence availability and quality whilst maintaining the integrity of the evidence throughout the digital examination process. One further principle is the documentation principle, which stated goal is to permanently record all information relevant to and/or generated during the digital examination.

Objectives cover the uniqueness of each forensic examination, which necessitates a non-checklist approach. They are aimed at practitioners to formulate examination goals and can greatly support the selection of steps necessary. By their very nature, objectives are (sub-) phase overarching and can help in event reconstruction.

Discussion

The analysis of the hierarchical, objectives-based framework for the digital investigations process in the context of this thesis yields insight in the granularity of examination steps and the connection between steps and overarching goals/principles.

The model represents a high-level practitioner's view according to [PBM08, pp. 116-117] and describes a forensic process divided into examination steps.

Ideas from the structuring of the phases, namely the use of data collection, data analysis and findings presentation are integrated into Sets of examination steps of the DCEA introduced in this thesis as described in Section 4.4.1. The incident response step is omitted because it somewhat blurs the boundary between forensics and incident response with the quick re-establishment of business continuity. But some ideas regarding the formulation of a strategy before the start of any examination is taken on board in the step of operational preparation as described in Section 4.4.1.2. Especially from the phase of incident closure, the disposal of evidence is integrated into the step of documentation as described in Section 4.4.1.6.

Sub-phases can be integrated into the investigative action, which is briefly addressed in the outlook of this thesis as part of an effort to formulate a digital forensics examination into a tuple (see Section 9.1 of this thesis). They are largely disregarded in DCEA because of complexity involving them and are left for future work.

We integrate the overarching principles of evidence preservation and documentation in the process accompanying documentation as described in Section 4.4.1.6.

The objectives as a motivator for the whole investigation are largely disregarded in this thesis because they receive only a brief mention in [BeC05, p. 149] and lack a definition. We further believe that these objectives are highly case-specific and can result in a biased examination, if the wrong objectives are formulated.

3.1.10 Defining Digital Forensic Examinations and Analysis Tools Using Abstraction Layers [Car03]

The objective of the model from [Car03, pp. 3-8] is to describe the purpose and goals of digital forensic analysis tools. This is mainly done in order to identify errors introduced by the tools and to provide requirements that the tool must follow. The raison d'être for this model are two fundamental problems for digital forensic tools identified by [Car03, pp. 2-3] as:

- Complexity Problem - acquired data often at the lowest and most raw format, requires great skill to interpret and inefficient for each expert to possess the skills,
- Quantity Problem - huge amounts of data to be analysed, inefficient to analyse each single piece, data reduction by grouping data into a larger event and/or removing known data,

both of which are only on a steady increase with ever more powerful CPU and parallel computing combined with steadily increasing mass storage space sizes. The author applies the concept of abstraction layers, i.e. are used to analyze large amounts of data in a more manageable format, to digital forensic tools. Thus, digital forensic tools can be described as a composition of abstraction layers to translate the data, often with a presentation functionality to show the results of the tool use in way that is intelligible to humans (i.e. the forensic expert). Abstraction layers in digital forensics following [Car03, p. 3] can be described as a function of inputs and outputs (see Figure 18).

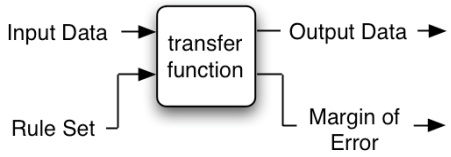

Figure 18: Abstraction layer inputs and outputs (modified from [Car03, p. 4])

Additional input is a rule set to describe how the input data should be processed. The abstraction layer produces output data and a margin of error. This margin of error addresses to different kind of errors [Car03, p. 5]:

- *Tool Implementation Error* (programming and design errors) and
- *Abstraction Error* (due to simplifications when generating the layer of abstraction

The tool implementation error is suggested be estimated based on the number of faults found in recent years and the severity of them. The margin of error posed by the abstraction errors is described in more detail and can be zero, representing a *Lossless Layer* (zero margin of Abstraction Error, e.g. file system). *Lossy Layers* have a margin of error greater than zero by employing a lossy compaction or compression and basically employ a surjective function, i.e. mapping a number of source inputs to a single output (e.g. IDS alerts only identifying known attack patterns). The author derives analysis tool requirements as:

- Usability - provision of data at a layer of abstraction and format clearly and accurately avoiding incorrect interpretations by the examiner,
- Comprehensive - identification of both inculpatory and exculpatory evidence, needs examiner's access to all data at a given Layer of Abstraction,
- Accuracy - to solve the error problem, tools must ensure that output data is accurate and a margin of error is calculated for appropriate interpretation by the examiner,
- Deterministic - must always produce the same output and margin of error when the same translation rule set and input is used,
- Verifiable - result verification both manually and by using a second and independent tool set.

[Car03, p. 8] further lists a number of recommended features of analysis tools:

- Read-only - not a necessity but highly recommended when the acquisition tool produced identical copies, which can be restored at any given time,
- Sanity Checks - all data can be used as input data at an abstraction layer, but only some outputs will be valid, presentation tools should conduct sanity checks on output data.

The requirements from [Car03, p.8] represent a subset of those provided by [Dar10, p. 60] and thus underline the importance of those demands on tools and processes used in forensics.

Discussion

The model represents a computer scientist's view according to [PBM08, pp. 116-117] and translation functions applied to digital data within digital forensic tools. Although not elaborating on uncertainty or the structuring of the input data of transfer functions, the analysis of the model presented in [Car03, pp. 3-8] in the context of this thesis yields numerous insights. The idea of transfer functions is picked up and generalised for the sets of methods for the forensic process described in Section 4.3 of this thesis. It also serves as one basis for the forensic data types and their layout described Section 4.2. The recording of known margins of error of both types is an important part of process accompanying documentation, which is discussed in Section 4.4.1 of this thesis. The connection between error and loss is not elaborated in [Car03, pp. 1-12], implicitly one can derive that loss is a result of error. In this thesis, the connection between loss, error is formalised in Section 4.1 together with the addition of uncertainty.

3.2 Loss, Error and Uncertainty

The determination of error and uncertainty is a common problem in science in general. Wherever measurements need to be taken in the physical world (which is also extensively used in digitised forensics), the resulting data is always accompanied by an *error* [HaD04, p. 152-1]. Typically, the true value of the measurement cannot be established. The error denotes the difference between the true value and the measured value [Rec06, p. 4]. An upper bound on the error represents (numerical) accuracy. In the context of measurements in the physical world, *uncertainty* represents an estimate (and therefore not a guarantee) of the error by quantifying the expected accuracy. This uncertainty of a measured result consists of several components, which can be differentiated into components of uncertainty arising from a *random effect* and components arising from a *systematic effect* [TaK94, p. 2]. Random effects as components for uncertainty can be addressed by using any valid statistical means treating data. Systemic effects, according to [TaK94, p. 2] can be evaluated based on scientific judgement, which may include:

- previous measurement data,
- experience with, or general knowledge of, the behaviour and property of relevant materials and instruments,
- manufacturer's specifications,
- data provided in calibration and other reports, and
- uncertainties assigned to reference data taken from handbooks.

Digitised forensic, of course, involves measurement in the physical world and their subsequent digitalisation. However, addressing the random or systematic effects of physical measurements on physical objects is outside the scope of this data-centric thesis. It should be obvious that for the use in digital and digitised forensics in a data-centric view, the concepts need to be adapted. First existing ideas are outlined in the remainder of this Section 3.2.

Of particular interest when trying to adapt measurement aspects to digital and digitised forensics seem to be the *systematic effects* component by evaluation based on scientific judgement. *Random effects* as part of the error component are likely to be relatively rare and very difficult to express in digital forensics. The evaluation aspects formulated in [TaK94, p. 2] can be translated into aspects relevant for digital and digitised forensics. In this context the evaluation of previous measurement data could be translated into the evaluation of previous forensic examinations, divided into examination steps (see Section 4.4). The experience with, or general knowledge of, the behaviour and property of relevant materials and instruments could be translated into the experience with, or general knowledge of, the behaviour and property of relevant materials and instruments as methods for the forensic process and forensically relevant data (see Sections 4.3 and 4.2). Manufacturer's specifications could also be translated to manufacturer's specifications of methods for the forensic process and forensically relevant data. Data provided in calibration and other reports could be translated into data provided by the methodical testing of forensic tools and methods. Uncertainties assigned to reference data taken from handbooks could be translated to known malfunctions in methods of the forensic process and the IT systems they work on (both target and examining systems).

One general concept, originally devised for traditional forensics but equally applicable to digital and digitised forensics is centred on the examiner. Forensic examinations, including examinations using IT systems, are in principle subject to error, uncertainty and loss during the location, acquisition, examination and analysis of traces. Beginning with the arrival at a crime scene (physical or virtual), as stated in [InR01, p. 196], *errors of omission* (i.e. where a person fails to do something they should) and *errors of commission* (i.e. where a person does something they should not) can occur. Mistakes made during the examination cannot be rectified and are magnified with each subsequent step.

In the following, an existing first informal attempt at examining loss, error and uncertainty is summarised from [Cas02, pp. 1-45] and examples from the scenarios researched in this thesis (Section 5, Section 6 and Section 7) are given to illustrate the problem, which is further discussed in detail in Section 4.1, where for the first time (to the best knowledge of the author of this thesis) formal definitions for loss, error and uncertainty for forensic examinations are proposed.

3.2.1 Loss in forensic examinations

The event reconstruction in forensic examinations is typically incomplete. This can be the result of *loss* of digital data containing evidence [Cas02, p. 2]. Digital data contained in IT systems can get deleted or overwritten during the normal operation of an IT system or as the result of malicious activity. Furthermore, such loss can occur by digital data not being located, acquired, examined and analysed or even destroyed during the forensic examination (most prominently during live forensics).

For the use case of the examination of main memory (see Section 5.2) such loss can include e.g. a software error in an operation system executing on the IT system used for video surveillance, which causes a system reboot, resulting in the loss of all volatile data. For the use case of hidden data in USB mass storage and device impersonation (see Section 6.2), such loss can include the failure to detect hidden storage areas and thus the exclusion of potential case-specific, forensically relevant data from the examination. Such failures can be attributed to the tools that are used to detect the real size of storage space in mass storage media. For the scenario of digital dactyloscopy (see Section 7.2) such loss can result from the failure to detect potential areas of traces on substrates by methods used for feature extraction and subsequent classification.

3.2.2 Error in forensic examinations

When electronic evidence is located, acquired, investigated and analysed, every aspect has some degree of *error* [Cas02, p. 2]. Errors are introduced into a forensic examination e.g. by:

(1) The IT system to be examined (hard- and software),
(2) The IT systems involved during the examination (hard- and software),
(3) The operators of the IT systems,
(4) The responders to a crime scene,
(5) The forensic examiners.

In the following, selected sources of errors are motivated using examples to highlight the validity of the claim from [Cas02, p. 2].

A selection of erroneous activity by first responders and IT system operators is highlighted in [Cas02, p. 32]; examiners should expect some *evidence dynamics*, i.e. any influence that changes, relocates, obscures, or obliterates evidence. This can originate from malicious (e.g. anti-forensics, see Section 2.6) or non-malicious activities (e.g. mistakes during intrusion response actions).

Addressing the forensic examiners (5), for instance, during the analysis part of a forensic examination, *reconstruction or interpretation errors* can occur. One reason, next to e.g. the complexity and the amounts of data (1), is evidence dynamics, which refers to influence on items of evidence that changes, relocates, obscures, or obliterates (physical) evidence, regardless of intent [ChT00, p. 52]. Evidence dynamics is relevant until the ultimate adjudication of a case. Although originally only applied to physical evidence, it can easily be argued, that evidence dynamics is also relevant to digital data containing evidence (e.g. when using SSD devices [BeB10, p. 11] or IT systems that are left in operation in the time between an incident and data acquisition or when performing live forensics).

For the use cases of the examination of main memory (see Section 5.2) and for hidden data in USB mass storage and device impersonation (see Section 6.2), errors could occur when the usage of carving tools (see e.g. [Ham18, pp. 188]) is considered, where due to the limited header/footer definition space a typically large number of false positives with regards to recovered documents/objects are common. For the use case of non-destructive latent fingerprint examination (see Section 7.2) as part of digitised forensics, errors could occur during feature extraction and subsequent classification tasks. We would attribute those errors to (2) and thus the systems involved in the examination and their configuration.

3.2.3 Uncertainty in forensic examinations

Uncertainties in forensic examinations can arise from a variety of reasons. Hard- and software used to locate, acquire, investigate and analyse the traces can behave in an erratic way that is not reproducible. Further, traces (and their correlation) may not be sufficient to derive enough information. This is additive to the fact, that electronic evidence gathered from an IT system can be tied directly to a person only on very rare occasions, e.g. when some sort of biometric authentication is added to the access credentials. Even then, it also comes with some degree of uncertainty attached to it because of the error rates inherent those biometric systems, such as the false acceptance rate or the false rejection rate (see [Vie06, p. 24]). This means that during the reconstruction of events based on associations (see Section 2.2) there will always be uncertainties [Cas02, pp. 2-4]. These uncertainties can be separated into *temporal uncertainty* and *uncertainty of origin*. Temporal uncertainty obstructs the determination of absolute time and the determination of the sequence of events when examining electronic evidence. Uncertainty of origin obstructs the determination of the source of electronic evidence [Cas02, pp. 6-11].

For the use case of the examination of main memory (see Section 5.2) such uncertainty can exist, for example the presence of artefacts with data containing hardware information pointing towards removable hardware. The mere presence of this data, especially if it represents non-essential data

(see Section 2.7.2), often leaves temporal uncertainty and uncertainty of origin. This uncertainty does not only apply to the host but also for the use case of hidden data in USB mass storage and device impersonation (see Section 6.2), such uncertainty can also affect data within processor controlled components acting as clients (both external and internal components).

For the scenario of digital dactyloscopy (see Section 7.2) such uncertainty can result from the application of image enhancement filter that overemphasise minor differences of texture to edges, the former sometimes caused due to varying lighting conditions the during acquisition.

Discussion

The ideas highlighted in [Cas02, pp. 1-45], most notably the notion of loss, error and uncertainty have to be applauded and pave the ground for the work introduced in this thesis. However, the terms error, uncertainty and loss are sometimes used in an informal way. We did not find means to integrate the computer scientist's view and concise methods to assign a layer of certainty.

By the time of writing this thesis, no universally accepted measure of error, uncertainty and loss in forensic examinations exists. Also, no formal definition of the terms error, uncertainty and loss for usage in forensic examination exists. In Section 4 we introduce a Data-Centric Examination Approach DCEA, whose main aim is to provide a qualitative measure for error and loss and the resulting uncertainty after proposing a formal definition of those terms in Section 4.1.

3.3 Forensics in different data streams contained or communicated in IT systems - a pre structuring problem discussion

As stated in [FHP+12, pp. 1-13], digital forensics is a multi-disciplinary science, involving several well-established research areas such as computer science, computer engineering and law. Several attempts to further specify digital forensics lead to a division into:

- *Computer forensics* (i.e. activities associated with the identification and preservation of computer or electronic evidence in support of some official or legal actions [New07, p. 17]),
- *Network forensics* (i.e. the use of scientifically proven techniques to collect, fuse, identify, examine, correlate, analyze, and document digital evidence from multiple, actively processing and transmitting digital sources for the purpose of uncovering facts related to the planned intent, or measured success of unauthorized activities meant to disrupt, corrupt, and or compromise system components as well as providing information to assist in response to or recovery from these activities. [Pal01, p. 27])
- *Media forensics* (i.e. analysing a digital asset, to give an assessment on such a content and to extract information that can be useful to address and support an examination linked to the scene represented in that specific digital document. [CAP+10, p. 131]).

However, when looking at digital forensics from a data-centric view of IT systems in digital and digitised forensics as proposed in this thesis, all of those aforementioned areas need to be considered. Media forensics is particularly relevant in digitised forensics (see Section 3.4 and Section 7) to assure authenticity and integrity of the digitalisation of physical trace evidence.

3.3.1 Selected overall topics in digital forensics relevant to all data streams

In digital forensic examinations a number of topics exist, which have a great influence on the quality and amount of data that can be acquired, investigated and analysed. All of the topics selected in this section need to be considered far in advance of any specific examination, placing them firmly into the examination step of strategic preparation as discussed in detail in Section 4.4.1.1.

3.3.1.1 Order of volatility

An important property that needs to be observed in digital forensics is the order of volatility of digital traces in IT systems. According to [FaV05, p. 6] different types of data residing in IT systems as digital traces have different life spans (see Table 2).

Data	Life Span
Registers, peripheral memory, caches, etc.	Nanoseconds
Main memory	Ten nanoseconds
Network state	Milliseconds
Running process	Seconds
Disk	Minutes
Floppies, backup media, etc.	Years
CD-ROMs, printouts, etc.	Tens of years

Table 2: The expected life span of data [FaV05, p. 6]

The life span of data and prioritises the data acquisition process according to the order of volatility. Because in digital forensics the extend of the incident at the start of an examination is typically not known, the order of volatility according to [FaV05] is always of the highest importance and provides the foundations of the data acquisition from an IT system. It influences heavily the choice of proceedings when notifying an incident or arriving at a suspected IT-related crime scene.

3.3.1.2 Post-mortem vs. Live digital forensics

Post-mortem forensics (see also Section 2.1.1) is a type of examination in digital forensics that can allow the inspection of contents (i.e. locating, acquisition, examination and analysis) of objects with a low level of volatility, typically mass storage devices (see also [CVD+11, p. 307] and Section 4.2.1.1). A post-mortem examination especially simplifies maintaining the security aspects of integrity and authenticity by creating forensic images (duplicates) of read-only connected mass storage devices and calculating a cryptographic hash sum using the most recent algorithm with no known collisions and retaining the original mass storage device in a chain of custody and only working on the forensic image. Thus, third parties can simply calculate the hash sum again and could detect integrity violations by a mismatch of those check sums. By retaining the physical mass storage device they can also prove the authenticity of the data contained in that image. However, for the generation of a forensic image the IT system has to be shut down (see also Section 3.3.1.3), involving the disconnection from networks, unmounting of encrypted discs, thus potentially causing a loss of a significant amount of potential trace evidence contained in objects with a high level of volatility (e.g. main memory, network state).

Live forensics (see also Section 2.1.1) is a type of examination in digital forensics that can allow the inspection (i.e. locating, acquiring, investigating and analysis) of the state of a running machine without disruption (see [CVD+11, p. 307] and Section 4.2.1.2 and Section 4.2.1.3). Important information can be derived from data gathered during live forensics (e.g. logged in users, active network connections, running processes). The downside of live forensics is the potential alteration of content in volatile and non-volatile objects, potentially lowering the integrity of the trace evidence.

3.3.1.3 Structural impact of forensic tools and procedures on the example of “pulling the plug”

The choice of forensic tools and procedures and the sequence of their application (see also Section 4) highly influence the amount of data that can be located, acquired, investigated and analysed, especially due to the *structural impact* of forensic tools and procedures. Using one technique can support or prohibit applying another technique. A very specific procedure is the removal of power and/or network connections of a running IT system, i.e. the execution of a post-mortem digital forensic examination (see Section 3.3.1.2). No general recommendations can be given, the decision to remove network connections or the power supply is inherently case-specific. It depends (especially when a suspected intruder is deemed to still be present on the target IT system) on the weighing of the importance of gaining potential valuable trace evidence

material against the potentially continued violation of the security aspects. The option of observing the attacker is also used in honey pots (see e.g. [JoS11, pp. 298-299]).

3.3.1.4 Forensic Timelining - The importance of time in forensic examinations

The event reconstruction requires the ordering of associations of in *space* and *time* (see Section 2.2). To order the associations with respect to *time*, forensic timelining is applied. In digital forensics it follows the concept that most digital data, e.g. if stored on mass-storage devices, carries a time stamp of some sort. Time is often maintained in battery backed-up hardware clocks (sometimes called CMOS clock) inside components of IT systems such as workstations, servers but also network infrastructure elements such as routers, switches etc. and storage solutions such as Network Attached Storage (NAS) devices. It is used for timestamps in file system activities (modify, access, create/change - see Section 4.2.1.1.2), log files etc., which in turn are relevant for forensic analysis of incidents in IT systems (see also [Glad04, pp. 24-33].

However, a number of problems are associated with timestamps and their use for forensic timelining. First of all, timestamps can easily be forged [Glad04, p. 30], e.g. by malicious manipulations of the system clock using physical access to the system or issuing of a command using the operating system (typically requiring administrator privileges). But also non-malicious properties can render timestamp-based information less reliable. In networked environments, the system clocks can be subject to a time skew. Even simpler, sometimes it is not known if a given timestamp was recorded using local time or if it was adapted using time zone information.

System operators can take measures to ensure correct time bases, such as using radio controlled clocks [Lom03, p. 1] or adjust the time using network time information such as NTP [Kos14]. These measures obviously have to be implemented prior to a suspected incident, typically by the system operator. Regarding the sets of examination steps used by the Data-Centric Examination Approach (DCEA) introduced in this thesis, such activities belong to the Strategic Preparation *SP* (see Section 4.4.1.1).

To overcome some of the problems with timelining, according to [Gla04, p. 30], two procedures have been suggested, time bounding and dynamic time analysis. In time bounding, as described in [Gla04, p. 31], a priori knowledge about certain events is exploited to estimate the time of events. If a given event is known to be followed by a second event, which itself is known to be followed by a third event, then the second event is bounded in time such that it cannot have happened before the first event and not after the third event. Knowing the details about the first and the third event allows to establish a time bounding for the second event. Dynamic time analysis, as described in [Wei02, p. 1], outlines am method to correlate data contained inside a file against the MAC times of that file in the file system. It can be used to calculate the offset between those two sources of timestamps and thus determine the real system time, provided that the clock sources for both the internal timestamp and the modify, access, change (MAC) timestamps and the clock pulses have no or only little drift [Gla04, p. 33].

3.3.1.5 Forensic duplicates and forensic imaging

Forensic investigations in the digital domain are special in the sense that the information can in theory be duplicated in all their digital properties. Contrary to other forensic disciplines (e.g. crime scene forensics), not a representation of actual analogous signals is sought, but their digitised representation. In digital forensics the particular sensor signal leading to a digital 1 or 0 is not evaluated (see also [BFGK09, pp. 97-98]). If we stick to the digital abstraction of the signal, digital evidence can be duplicated and a copy can be examined as if it was the original. Working with a copy is common practice to avoid the risk of altering or damaging the original evidence (see [Cas11, p. 26]). However, it also means that other assumptions about the trustworthiness have to be placed. Typically, cryptographic means are employed to ensure the verifiability of maintaining the security aspect of integrity during the acquisition of the digital signals and thereafter, which is also to be documented typically in the chain of custody (see Section 2.4). In this thesis we discuss the documentation step in Section 4.4.1.6 in depth.

Generally, the usage of a *forensic duplicate* that behaves identical to the original in the digital domain is beneficial e.g. due to structural impact of forensic tools (see Section 3.3.1.3 or human error (see Section 3.2.2).

In mass storage forensics (see also Section 4.2.1.1), *forensic imaging* allows for a one to one image for further investigation/analysis. This technique is often used in post mortem forensics (see Section 3.3.1.2). It is preferable if the mass storage device can be acquired as a whole (i.e. a sector-wise collection of all addressable memory space of the device). If the storage capacity of the system selected to contain the image is smaller than the source device or other crucial factors such as time constraints inhibit the creation of full disk images (known as the *volume challenge*) *partial images* have been suggested (see e.g. [GrR15, pp. S34-S35]). Integrity-ensuring mechanisms (typically cryptographically computed hash sums deemed to be secure for the duration of proceedings and onwards) are employed to prove that no alterations to the source and the resulting images exist. Only in mass storage forensics it is conceivable to show that fact due to the low volatility of data on mass storage devices (see Section 3.3.1.1).

(Partial) Forensic imaging is also used in main memory forensics (see Section 4.2.1.2). Here either hardware-based acquisition devices (e.g. [CaG04, pp. 56-60]) can be used or software-based acquisition tools can be employed. The former requires this hardware device to be present and active prior to an anticipated event (which we look into as strategic preparation in Section 4.4.1.1) while the latter alters the very source it is about to acquire. However, immediately after the collection of that memory data at the latest, integrity-ensuring mechanisms need to be employed to prove that no further alteration took place after the image file entered the custody of the examiner.

(Partial) Forensic imaging is also used in network forensics. Here, the network data is recorded at the lowest possible layer (see Section 4.2.1.3.1) typically into a file representing the network packets (see e.g. [KHA+10, p. 99]). In [KHA+10, p. 97] we also add mechanisms to assure authenticity for the forensic network stream data collection tool named *Linux Transparent Forensic Bridge (LFTB)*.

Forensic imaging fails, however, in situations where only logical access to storage space is available (e.g. in cloud storage environments). However, even in this situation, the acquired objects (e.g. files, folders etc.) are copied whilst maintaining the integrity of the collected data, ideally together with meta data (e.g. access rights, MAC times etc., see Section 4.2.1.1.2) gathered using the tools of the operating environment used to copy the data.

3.3.1.6 Forensic sound deletion (sanitisation)

The storage space chosen to contain the forensic images and forensic duplicates (see Section 3.3.1.5 needs to be sanitised by means of a *forensic sound deletion* (see also [Cas11, p. 212-214]. The rationale behind that is to avoid any cross-contamination with previous evidence and other material that is not case-related. Both could be sources of error as outlined in Section 3.2.2.

This sanitisation is best performed during strategic preparation (see Section 4.4.1.1) due to the potential long duration of the forensic sound deletion. Every user accessible sector (see also Section 6) is cleared by a known sequence of data. For maximal transparency we suggest to remove any potential alterations performed to the drive configuration (see Section 4.2.1.1.2) such as Host Protected Areas (HPA) or Drive Configuration Overlays (DCO) prior to use as forensic storage. We argue for a sequence of zeros (see Listing 1).

```
sudo dcfldd if=/dev/zero of=/dev/sdX
```

Listing 1: Example of a sanitisation in a unix-like environment using a pattern of zeros

Here we assume a UNIX-like environment (e.g. the CAINE [Bas19] bootable forensic Linux distribution). The drive description that is set as the variable *X* in sd*X* is situation-specific and has to

be identified after including the storage media into the forensic environment. Great care needs to be taken to ensure the correct drive description is entered as all information contained in that drive is irretrievably lost in the forensic sound deletion process.

Another benefit of the forensic sound deletion of forensic storage media is an accessibility check of all sectors and thus the whole user addressable portion of the drive. Errors shown during the sanitation process (apart from errors writing past the user addressable area) should lead to an immediate exclusion of that drive for further forensic use.

Afterwards, the success of the sanitisation is easily checked in said forensic environment using the line from Listing 2 below).

```
sudo cmp /dev/sdX /dev/zero
```

Listing 2: Example of a blank check of sanitised media using a unix-like environment

This line performs a blank check and terminates with an error of sectors past the user addressable area if the forensic storage space is fit for purpose. Again, the drive description that is set as the variable *X* in sd*X* is situation-specific and has to be identified after including the storage media into the forensic environment.

Obviously, other patterns could be used, indeed are suggested by [Gut19]. These suggestions, apart from controversy regarding its prolonged applicability, cover an attacker model where an attacker deliberately tries to recover previous drive content using means not provided by the drive itself (e.g. electron microscopy). This attacker model is not applicable to our setting, as we only want to rule out error and non-malicious activity in that part of forensic storage sanitisation.

After looking into selected general topics that can be applied to all data streams we now look into selected specifics of each data stream based on their structure of data organisation, selected challenges and characteristics that can be used to gather information obtained in IT systems.

3.4 Selected aspects of the state of the art in digitised forensics

Digitised forensics is a relatively new branch of forensics (see [Sri10, p. 40] describing the very similar field of computational forensics). The idea of digitised forensics is to detect and acquire crime scene traces digitally and (preferably) contactless. This helps traces concurrency, whereby the examination of the trace using chemical or physical ages renders the trace inaccessible to other examinations, e.g. treating a fingerprint trace with cyanoacrylate to reveal its morphology prevents the same fingerprint trace examined for remnants of illegal drugs. If, however, contactless equipment is used to reveal the morphology of the trace, the same trace could be re-examined to find answers to other questions.

Generally, crime scene forensics is based on the principles of transfer of traits and matter and the principle of the divisibility of matter [InR01, p. 77] from Section 2.2.1 and differentiates between *patent traces* (visible with the naked eye), *latent traces* (hidden or unseen by the naked eye) and *plastic impressions* (imprints into pliable substances acting as substrates) as shown on the example of fingerprints in [YaF11, pp. 7-3 - 7-4]. Especially latent traces need to be developed [YaF11, p. 7-2], i.e. some agent has to be applied to render the trace visible. But often in doing so, renders other features of the trace inaccessible (see above).

(Partial) digitalisation is achieved or actively researched on other trace types apart from fingerprints and includes firearm and toolmarks/ballistics (see e.g. [FVH+13, pp. 86650F-86650F-12]), fibre traces (see e.g. [AKD+15, pp. 91-96]) and locksmith forensics (see e.g. [KCD+13, pp. 367-379]) amongst others.

If contactless means of trace localisation, acquisition, investigation and analysis (sort of 'digitalised development') are applied to latent (i.e. invisible) traces, persuading a judge/jury and assuring a process has been conducted flawless and meets the required standards for evidence (see

Section 2.7) becomes a whole new challenge with little or no visible clues on the substrate carrying the trace.

Generally, digital forensics seeks a similar goal as biometrics does (see Section 2.8), i.e. the identification and verification of the presence of people at the crime scene - at some point in time, that is, before or during the suspected incident. Ongoing research is the age estimation of some trace types (see e.g. [Mer14, pp. 1-214] for fingerprints), which, if successful, could dramatically reduce the number of suspects for an incident.

However, as remarked by the reviewer Eoghan Casey, profound differences between biometric user authentication and forensic science processes exist, where the latter entails an entirely different reasoning such as the Bayesian probabilistic approach (see e.g. [Eur20, pp. 15-18]).

Important differences between biometrics for user authentication (see e.g. [Vie06, p. 1]) and digitised crime scene forensics (or digitised forensics SF for short) are also that the pattern recognition pipeline is run trough on latent residue/impressions etc. and not a person being present and consenting to the acquisition of behavioural/physiological traits and the output of that pipeline is an annotated trace and substrate picture for assessment by the forensic expert (see also [KHD+12, p. 1507]). The following Figure 19 shows the general process:

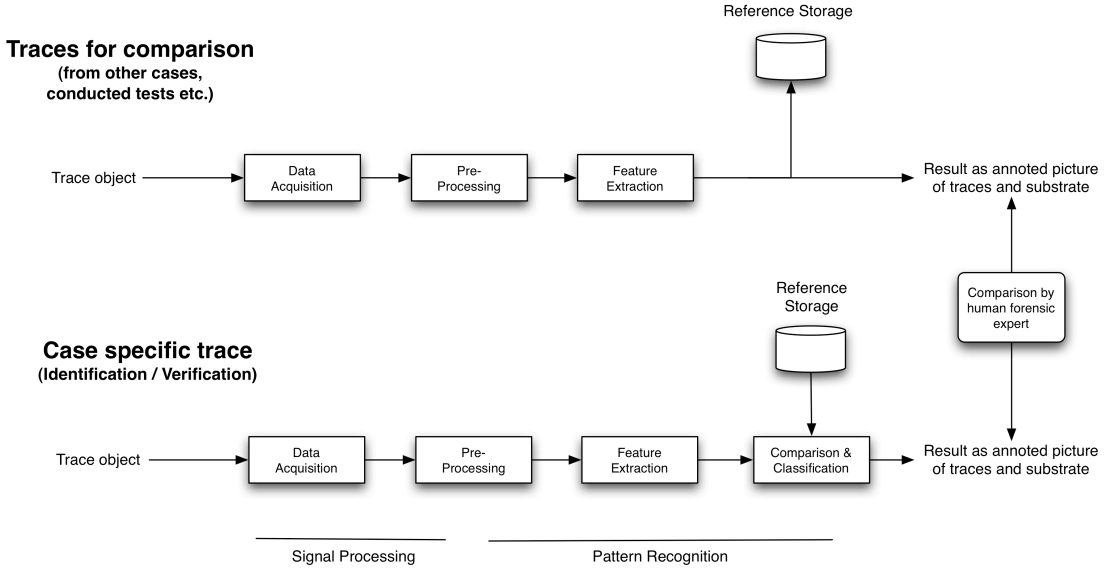

Figure 19: Digitised forensics as a comparison process involving signal processing and pattern recognition based on [Vie06, pp. 19-20]

The pattern recognition pipeline can also be run trough to only solve very specific subtasks (e.g. separation of trace and substrate as 'digital lifting' of traces). Recognised features (e.g. minutia points in latent fingerprints) can be used to automatically annotate the fingerprint images. In some cases (e.g. in firearm and toolmarks/ballistics) even some automated matching between feature sets of two or more specimens can be executed (as depicted in Figure 19).
However, it needs to be stated that, as of now, no final decision is reached by a computing device, all results have to be presented to trained forensic examiners, which have to reach the final conclusion of agreement between two traces, disagreement between two traces or inability to reach a conclusion (e.g. due to lack of detail or features). In the following, we will look into one methodology specifically designed to aid forensic examiners in the field of fingerprint verification.

3.4.1 The ACE-V methodology for fingerprint examination

The ACE-V methodology as summarised by [Van11, pp. 9-12 - 9-14] describes a process in the investigation of latent fingerprint images by a forensic examiner. Generally, the examiner com-

pares an image of a processed latent fingerprint of an (yet) unknown person against a fingerprint image of a second fingerprint of known origin to establish whether they are agreement or disagreement.

According to [Van11, p. 9-12] it gives the examiner specific phases of examination that can be used to document the perception, information- gathering, comparison, and decision-making that take place during an examination of fingerprints. It consists of the following phases [Van11, pp. 9-13 - 9-17]:

- Analysis: assessment of a fingerprint as it appears on the substrate, uses analysis of clarity to establish the levels of detail that are available to compare, examiner determines whether the fingerprint is sufficient for comparison,
- Comparison: direct or side-by-side comparison of friction ridge details in two prints are in agreement based upon similarity, sequence, and spatial relationship, comparative measurements of first, second, and third level details along with comparisons of the sequences and configurations of ridge paths,
- Evaluation: final determination by the examiner as to whether a finding of individualization, or same source of origin, can be made
- Verification: independent examination by another examiner re-running ACE resulting in the same conclusion.

This process is decidedly not an exclusive process of digitised forensics; it could in theory be conducted without IT systems and always requires a human forensic fingerprint expert as examiner.

3.5 Summary of selected aspects of the state of the art in digital and digitised forensics

In this Section we researched the main starting points of the Data-Centric Examination Approach DCEA introduced in the following Section 4. For this outlined and discussed selected aspects the state of the art of selected aspects in digital and digitised forensics.

The starting point was set by the review of existing models of the forensic process to cover digital forensics. We looked into challenges solved by the respective models but also identified remaining gaps. Some of the latter it is that many models ([Dar10], [KCG+12], [New07], [KH02], [Pol08], [FrS07], [CaS03], [RGM+06], [BeC05]) often address only the high level practitioner's view from [PBM08, pp. 116-117]. However, each of those models yielded some input for the content of the sets of examination steps of the DCEA approach (see Section 4.4).

We have looked into loss, error and uncertainty briefly in general and more detailed in the context of digital forensics. First attempts to tackle this problem, which are to be highly applauded, still have the gap that no formal definition based on the data containing information in IT systems is given. The existing information, however, forms a basis for our modelling of loss, error and uncertainty as part of the Data-Centric Examination Approach DCEA based on the data containing information in IT systems D_I that is introduced in Section 4.1

Further, we researched several overarching aspects in digital forensics including the order of volatility, properties of post mortem vs. live digital forensics, the structural impact of forensic proceedings, the importance of time in forensic examinations, forensic imaging, forensic sound deletion. Each of those overarching aspects have a direct influence in the modelling of the three pillars of the Data-Centric Examination Approach DCEA in the shape of the forensic data types (Section 4.2), sets of methods for the forensic process (Section 4.4) and sets of examination steps for the forensic process (Section 4.4).

For the fairly new discipline of digitised forensics, we researched properties regarding the forensic process itself. We showed on an example as to how that fits into accepted proceedings of crime scene forensics. Our findings in the modelling of the three pillars of the Data-Centric Examination Approach DCEA in the shape of the forensic data types (Section 4.2), sets of methods for the forensic process (Section 4.4) and sets of examination steps for the forensic process (Section 4.4) regarding digitised forensics.

4. A new Data-Centric Examination Approach (DCEA) for the forensic process in digital forensics and in digitised forensics

The contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Mario Hildebrandt, Tobias Hoppe, Robert Altschaffel, Claus Vielhauer, Carsten Schulz, Ina Grossmann, Michael Ulrich, Sebastian Breß (in descending order according to the occurrence of co-authorship): [KHD09], [KDV12], [KDV15], [KHA+10], [AKD09], [KHA+09], [KLD+11], [HML+11], [HKD09], [HKG+11], [HKD11a], [BKS13]

From Section 2 we use the terms and definitions laid out for the remainder of the thesis. The properties of forensic examinations described therein are used for the modelling of the process in digitised forensics, namely the forensic data types (Section 4.2.2) and the sets of methods of the forensic process (Section 4.3.2). In Section 2 we outline objectives of forensic examinations that go a lot further than just court proceedings and use them in the modelling of the sets of examination steps for both digital and digitised forensics (Section 4.4). Naturally, the description of the chain of custody features importantly and finds its use in the modelling of the sets of examination steps for both digital and digitised forensics (Section 4.4). Our discussion about finding a common language for incidents and their forensic examination steps also contributes to the findings in that Section 4.4. The description of Anti-Forensics influences our view on errors in forensic examinations in Section 3.2 and thus has an impact on modelling our understanding of error in relation to the data contained in IT systems from the perspective of forensic examinations (Section 4.1). Although positively only concerned with the technical and procedural implications of legal and data protection requirements, since we are no legal experts by any stretch of the imagination, we derived properties of our description of the forensic process that found its way into the shaping of the formalisation of loss, error and uncertainty (Section 4.1) and the forensic data types for both digital and digitised forensics (Section 4.2). Our description of our view on the biometric pipeline is used in the design of our view of the forensic process in digitised forensics, namely its forensic data types (Section 4.2.2).

We further use the findings of our survey into the state of the art in digital and digitised forensics described in Section 3. Namely, the layout and content description of the sets of examination steps (Section 4.4) is based on the findings of the review of existing models of the forensics covered in that Section 3. The existing ideas regarding loss, error and uncertainty outlined in Section 3 form the basis of our modelling thereof, based in the data containing information D_I in IT systems that is introduced in Section 4.1. Overarching properties of digital forensics based on a selection of the state of the art in digital and digitised forensics from Section 3 are used for the modelling of the forensic data types (Section 4.2), sets of methods for the forensic process (Section 4.4) and sets of examination steps for the forensic process (Section 4.4).

In this section a new Data-Centric Examination Approach (DCEA) is introduced, which in its underlying principles is suited for both digital and digitised forensics. We believe in the absence of a sensible argument that would discourage us to pursue the course of research introduced and justified in the following. However, it needs to be stated that the work concerning the application of the process model to digitised forensics is in its early stages of research and should be seen as a general proof of concept. We acknowledge, as suggested by the reviewer Eoghan Casey, the complexity related to digital evidence and the need for further study in this area (see also Section 9 concerned with future work).

At first we look formally into the data containing information that is gathered, stored, processed and communicated in IT systems to define loss, error and uncertainty in forensic examinations (data-centric view). Our research leads us to the contribution C1 from Section 1.3 of this thesis.

C1:

Formal description of loss, error and uncertainty using distinct sets of information contained in data (distinction into data containing information D_I, data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC} with only the latter solving specific incidents).

To achieve this, we use the method M1 from our methodology outlined in Section 1.2.

M1:

We describe of error, loss, and uncertainty regarding data contained in IT systems based on a modelling of the relationship of all data ever available, data used in all forensic investigations ever and the case-specific data for a given incident.

Following that we use and extend aspects of the model from [KHD09, pp. 2-3], which is also used in [KHA+09] and in the "Guidelines for IT Forensics" of the Federal Office for Information Security in Germany [BSI20a]. It is first proposed to a German speaking audience in [KHA+09, p. 478-480] and to an English speaking audience in [KHD09, p.1-3] and summarised and refined in [KDV15, p.87-88]. Our research regarding forensic data types leads us to C2 (incl. 2.1 and 2.2) from Section 1.3 of this thesis.

C2:

Establishment of a data-centric (data-driven) view in digital and digitised forensics by applying a structured layering of data based on selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 621] in various IT systems (from embedded systems/IoT to data centers) to formulate forensic data types:

C2.1 Adding semantics to data in the forensic context and establishment of 8 data types for digital forensics for selected use cases

C2.2 Adding semantics to data in the forensic context and establishment of 10 data types for digitised forensics for selected use cases

To achieve this, we use the method M2 from our methodology outlined in Section 1.2.

M2:

We apply the selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 621] to data stored, processed, communicated in IT systems to distinguish forensic data types for digital and digitised forensics. To achieve this, we construct layers of data by giving them semantics that support the forensic process. In further conjunction with the ISO/OSI model, we use a layering that is not mutual exclusive.

Following that we use and extend aspects of the model from [KHD09, pp. 1-3]. Our research regarding sets of methods for the forensic process leads us to C3 (incl. 3.1 and 3.2) from Section 1.3 of this thesis.

C3:

Establishment of a hierarchical mutual exclusive categorisation of methods for the forensic process:

C3.1 Distinction into 6 distinct classes based on the likeliness of availability for digital forensics for selected use cases

C3.2 Distinction into 10 classes based on the pipeline of the biometric process for a use case

To achieve this, we use the method M3 from our methodology outlined in Section 1.2.

M3:

We use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, pp. 3-4]. We further use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics.

Following that we use and extend aspects of the model from [KHD09, pp. 1-3]. Our research regarding sets of examination steps for the forensic process leads us to C3 (incl. 3.1 and 3.2) from Section 1.3 of this thesis.

C4. Establishment of 6 sets of examination steps for selected use cases:

C4.1 Digital forensics process specific properties of the examination steps based on a systematic review of existing models and selection of best fitting for a data-centric approach

C4.2 Digitised forensics process specific properties of the examination steps based on the application and adaption of the steps from digital forensics

To achieve this, we also use the method M3 from our methodology outlined in Section 1.2.

M3:

We use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, pp. 3-4]. We further use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics.

Following that we combine all contributions achieved thus far. Our combination effort leads us to C5 from Section 1.3 of this thesis.

C5:

Case-specific qualification of loss, error and uncertainty; and their representation based on forensic data types, methods of the forensic process and examination steps.

To achieve this, we also use the method M4 from our methodology outlined in Section 1.2.

> M4:
>
> We use forensic data types, sets of methods and sets of examination steps to provide a qualitative estimation on loss, error and uncertainty.

In the following Sections we will work our way around the methodology to achieve the contributions from above and start with the modelling of loss, error and uncertainty in forensic examination based on the data stored and processed inside of IT systems and communicated outside those IT systems.

4.1 Loss, error, and uncertainty in digital and digitised forensics with the Data-Centric Examination Approach (DCEA)

Ideally, the reliability of computer generated digital data containing electronic evidence could be based on the reliability of IT system and the process that generated the digital data (see [Cas02, p.2]). However, the assessment of the reliability of IT systems is notoriously difficult, since programmers are fallible and therefore implement errors in their applications (unintentionally or purposeful), which include the operating system, firmware of the main board and or plug-in cards and IT applications etc. but also forensic applications (see also [Cas02, p.3]). Furthermore, not only the implementation, but also all the other vulnerabilities from the incident taxonomy [HoL12, pp. 6-17] enhanced in [KLD08, p. 414] outlined in Section 2.5.1, namely the design, the configuration but also social engineering, add to the reduction in reliability.

In this context, *IT system use* has to be specified as well. For the remainder of this thesis, *normal use of an IT system* is defined as using the system as intended by the hardware/software specification. *Abnormal use of an IT system* thus represents system use contrary to what is intended by the hardware/software specification. This can be further separated into *intentional* and *unintentional abnormal use of an IT system.* The *intentional abnormal use of an IT system* constitutes an IT security incident and the *unintentional abnormal use of an IT system* is caused by an operator mistake.

The overarching goal pursued in this thesis is to find description and a first, universally applicable, qualitative measure for error, uncertainty and loss in forensic examinations. With the same starting point in digitised forensics but directed at probabilistic descriptions and including the digitalisation process itself, related work can be found in [Hil20]. The amount of data contained in an IT system cannot be determined fully before, during or after an incident as some data is inaccessible (e.g. CPU, MMU registers, data for internal use only in processor-controlled components, see Section 6). This rules out a percentage gauge type scale for now, instead qualitative measure is sought after. Thus, we do not consider probabilistic results, although this is already part of active research for some application fields [Hil20]. This is further motivated by the concept of three different categories of knowledge (or lack of) as modified by [For14, pp. 6-8] from a presentation given by former Secretary of Defence Donald Rumsfeld as follows:

- There are things that we *know that we know*,
- There are *known unknowns* - things we know that we do not know them,
- There are also *unknown unknowns* - things we do not know that we do not know.

Obviously, on practical terms, only the first two items can be addressed using existing techniques and knowledge. In the following we assume, with regards to data contained in IT systems, that also the unknown unknowns are included.

In the following we will use the item M1 of our methodology outlined in Section 1.2:

> M1:
>
> We describe loss, error and uncertainty regarding data contained in IT systems based on a modelling of the relationship of all data ever available, data used in all forensic investigations ever and the case-specific data for a given incident.

The notion of case-relevancy needs a clarification, thus we define *case-specific data* for the remainder of this thesis as follows:

> Case-specific data:
>
> Data contained in an IT system from single component up to networks of networks of systems employing a Random Access Stored Program architecture [Har71, pp.234-240] that can be a source of information leading to a successful event reconstruction of a suspected incident (on the target, victim and intermediate systems).

In order to provide a proposal for formal definition of error, uncertainty and loss in digital forensics DF and digitised forensics SF the amount of data contained in IT systems needs to be formalised. For that, the definitions of data and information are picked up from Section 2.1, which themselves are part of the Data-Information-Knowledge-Wisdom Hierarchy (DIKW) as provided in [Row07, p. 166] citing [Ack89]:

- Data are defined as symbols that represent properties of objects, events and their environment. They are the products of observation but are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural.
- Information is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate, store, retrieve and process data. Information is inferred from data.

According to those definitions the information is contained in data and data contains information. Following a top-down view incorporated in the Data-Centric Examination Approach (DCEA), a subset of the information contained in the universe is inferred from the data contained in an IT system, abstracting the way IT systems work, as depicted by the following Figure 20:

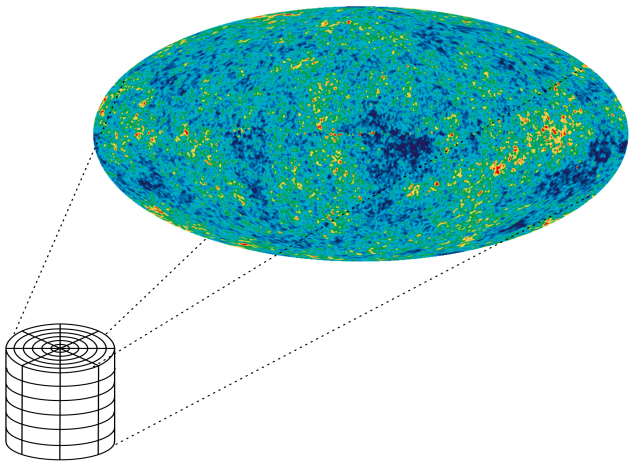

Figure 20: Subset of information contained as data in IT systems inferred from the universe as the sum of all information (depicted as cosmic microwave background from [Wol20])

In this figure a barrel-type representation of the data contained in IT systems is used, which is outlined in detail in Section 4.2. Following a data-centric approach containing information, as suggested by [Pol08, pp. 20-23] and outlined in Section 3.1.5, it can be argued, that data contains information, forming the set D_I. It can be inferred further, that in general forensically relevant data contains forensically relevant information, forming the set D_{IF} of all forensic examinations ever conducted. By following that statement, case-specific, forensically relevant data contains case-specific, forensically relevant information, forming the set D_{IFC} of one particular case. We argue for the possibility to establish set relations as shown in the following as:

$$D_{IFC} \subset D_{IF} \subset D_I \qquad (4.1)$$

The relationship is also illustrated by the following Venn diagram (Figure 21).

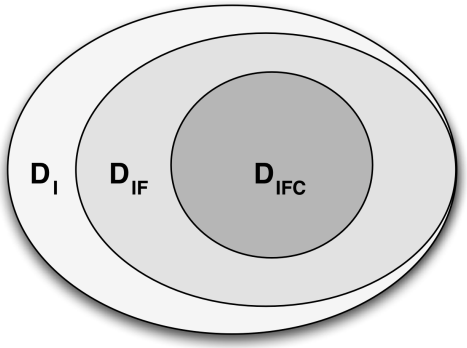

Figure 21: Relationship between the set of data containing information D_I, the set of data containing forensically relevant information D_{IF} and the set of data containing case-specific, forensically relevant data D_{IFC}

Thanks to the remark of the thesis reviewer Sabah Jassim during the Ph.D. defence we abstain from the usage of the term cardinality with regards to its strict formal definition. We assume the sets to be finite for the remainder of this thesis. Proving this assumption poses a very interesting subject that is beyond the confines of this thesis and thus is commended to future work.

The correctness of that relationship from (4.1) is shown by first looking for elements that are contained in the set of D_{IF} but not in D_{IFC}. One trivial example would be an isolated IT system that is not networked. Clearly, any forensically relevant data containing forensically relevant information based on network activity in this case is not contained in D_{IFC}. However, since by definition the D_{IF} contains all D_{IFC}, this also applies to the above example. So D_{IFC} forms a proper subset of D_{IF}.

D_{IF} is modelled as a subset of D_I on the grounds of the thesis from [InR01, p. 92] stating that the number of things that can be evidence is limited to the things that exist in the physical universe - in other words, anything can be evidence. Since this thesis is decidedly concerned with data contained in IT systems, any type of data therein can potentially be traces of evidence. However, a few exceptions can be named that justify a differentiation in the digital domain between D_I and D_{IF} with regards to practical considerations. Consider, for example, an electronic calculator that is switched off and only operates on fixed calculation routines. Clearly, this device uses digital data. However, that data is instantly erased after switching off the device or overwritten with the next calculation and thus has no considerable evidentiary value (see Section 2.7.2).

Data containing forensically relevant information form both the target D_{IT} and the result D_{IR} when using forensic tools and methods.

$$D_{IT}, D_{IR} \subset D_I \qquad (4.2)$$

Equation 4.2 contains this relationship and connects it to equation 4.1 where data from the investigation target D_{IT} and data from the investigation result D_{IR} can contain data that is forensically irrelevant (recall the calculator example) or from data that is relevant to forensic investigations. In

ideal circumstances the final result of the whole examination contains only data that is case-specific. The forensic examination strives to remove all case-irrelevant data in the course of proceedings.

A transfer function $f_{DF,\ SF}$ (equation 4.3) from the set of transfer functions $F_{DF,\ SF}$ for both digital forensics DF and digitised forensics SF (see also [Car03, p. 4] and Section 4.3 for its practical adaptation) translates data from investigation target D_{IT} into data of the investigation result D_{IR}:

$$f_{DF,SF}(d_{IT}) : D_{IT} \rightarrow D_{IR} \qquad (4.3)$$

The translation process can result in a data reduction (e.g. by filtering mechanisms) but also to a data increase (e.g. by mechanisms known to produce false positives such as file carving [Lau13], p. 431). Further, different transfer functions can return different data of the investigation result D_{IR} on the same data of the investigation target D_{IT}, which is for example revealed by the computer forensic tool testing programme CFTT [Gray12, pp. 1-6].

We thank the reviewer Sabah Jassim for his comments during the Ph.D. defence suggesting the use of a relation instead of a function. For practical purposes with working on existing datasets and to maintain the origin from [Car03, p.4], for this thesis we continue to use the term function whilst appreciating the suggestion of using relations. We commend a detailed discussion on this highly interesting subject to future work.

Also an inverse function $f_{DF,\ SF}^{-1}$ (equation 4.4) can exist in theory (e.g. for bitwise acquisition tools), which translates the data from the investigation result D_{IR} into data of the investigation target D_{IT}:

$$f_{DF,SF}(d_{IT})^{-1} : D_{IR} \rightarrow D_{IT} \qquad (4.4)$$

Equation 4.4 can also be of relevance for both sanity checks and forensic tool testing (see e.g. [Gray12, pp. 1-6] or [HML+11, pp. 1-6]). We will pick up the transfer function (equation 4.3) as the basis for the modelling for the sets of methods for the forensic process in Section 4.3.

Formally it can be stated for an idealised forensic examination that the data of investigation target D_{IT} consists only of case-specific, forensically relevant data containing case-specific, forensically relevant information:

$$\forall\, d_{IT} \in D_{IT} : d_{IT} \in D_{IFC} \wedge d_{IT} \notin D_{IF} \setminus D_{IFC} \wedge d_{IT} \notin D_I \setminus D_{IFC} \quad (4.5)$$

Equation 4.5 embodies that the demands for working on all case-specific, forensically relevant data (as containers for inculpatory and exculpatory evidence) on one hand and on the other the protection of person-related data and the privacy are met (see also our additional details in the appendix in Section 10.1.4). This includes that only case-specific, forensically relevant data is used and only for the original intention and according to necessity and proportionality.

Further, for an idealised forensic examination data from the investigation result D_{IR} consists only of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC}:

$$\forall\, d_{IR} \in D_{IR} : d_{IR} \in D_{IFC} \wedge d_{IR} \notin D_{IF} \setminus D_{IFC} \wedge d_{IR} \notin D_I \setminus D_{IFC} \quad (4.6)$$

Again, equation 4.6 reflects the demand for a strict case relevancy for the particular forensic investigation. Note, however, that nothing is said about the completeness of the case-specific, forensically relevant data. Case-specific, forensically data could have been missed, leading to false event reconstructions. Therefore, assuring that only case-specific, forensically relevant data has be located, gathered, investigated and analysed, on its own, is a necessary but not satisfactory condition. The Data-Centric Examination Approach (DCEA) introduced in this thesis based on forensic data types, sets of methods for the forensic process and sets of examination steps (based on and extending [KHD09, pp. 1-3]) is a first step towards addressing the latter by providing a qualitative measure for loss, error and uncertainty in forensic examinations.

As of yet, no assertion about the set relations of the data forming the investigation target D_{IT} and the investigation result D_{IR} regarding each element containing the case-specific, forensically relevant data has been made. To address this relationship, in the following definitions for loss, error and uncertainty are proposed together with explanatory examples. As we will see, error and uncertainty can lead to results leaving the set boundaries of D_{IFC}.

4.1.1 Loss in digital and digitised forensics

To define loss in digital and digitised forensics, each element of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} of the data of the target of the investigation D_{IT} and the data of investigation result D_{IR} and an element $f_{DF,\,SF}$ of the set of transfer functions $F_{DF,\,SF}$ need to be observed. This leads to the following definition:

Loss occurs if for a given element the amount of the case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} in the data of the investigation result D_{IR} is smaller than the data of the investigation target D_{IT} as a result of a transfer function $f_{DF,\,SF}$ returning no result.

Here we assume that all transfer functions terminate and the resources (e.g. storage space, time etc.) for it completion are unbounded. Clearly, this is a formal, idealised view. This definition is equivalent to the following formalisation (equation 4.7):

$$\forall\, d_{IR} \in D_{IT} \wedge d_{IT} \notin D_I \setminus D_{IFC} \wedge \forall d_{IT} \notin D_I \setminus D_{IFC} : \exists\, d_{IT} \notin D_{IR} \quad (4.7)$$

resulting in equation 4.8:

$$\exists\, d_{IT} \in D_{IT} \wedge d_{IT} \notin D_I \setminus D_{IFC} : f_{DF,SF}(d_{IT}) = \{\} \quad (4.8)$$

In the case of loss, no investigation result can occur, which is outside the set of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} as shown in equation 4.9:

$$\exists\, d_{IT} \in D_{IT} \wedge d_{IT} \in D_I \setminus D_{IFC} : f_{DF,SF}(d_{IT}) = \{\} \quad (4.9)$$

The resulting Venn diagram is depicted in the following Figure 22, where we define D_{IR} as a proper subset to D_{IT} and thank Sabah Jassim for this input of that nature during the Ph.D. defence.

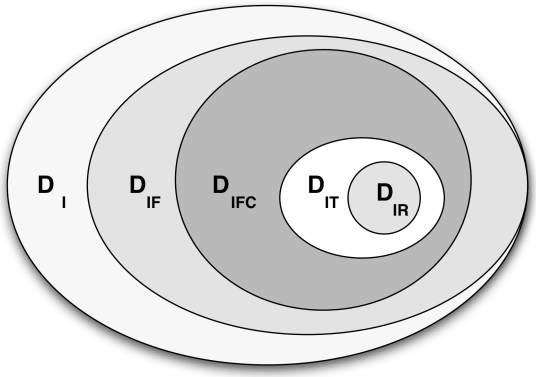

Figure 22: Relationship of the data of the investigation target D_{IT} and the data of the investigation result D_{IR} in the case of *loss*

Loss is a violation of the security aspect of availability of forensically relevant data contained in IT systems caused by normal and abnormal use of an IT system through the vulnerabilities of design, implementation, configuration and social engineering according to the incident taxonomy [HoL12, pp. 6-17] enhanced in [KLD08, p. 414] outlined in Section 2.5.1. This includes both the target IT system and the IT system used to examine an incident.

In digital forensics loss for example could result from overstressed resources inside an IT system. As stated in [Deri04, p. 2], the network acquisition tool wireshark[1] is known to drop network packets during high network load, resulting in an incomplete network traffic recording and thus loss in the investigation result (acquired network packages) compared to the investigation target (network stream).

An example for loss in digitised forensics is a wrongly chosen resolution (due to the choice of sensors or resolution settings) when acquiring latent fingerprints from the substrate. Case-specific, forensically relevant information contained in case-specific, forensically relevant data (e.g. Level 3 features such a pores, see [JCD07, p. 15]), which are present in the fingerprint residue (investigation target) are not acquired when using resolutions below 500ppi [JCD07, p. 15]), thus reducing the amount of Case-specific, forensically relevant information contained in case-specific, forensically relevant data in the investigation result.

4.1.2 Error in digital and digitised forensics

To define error in digital and digitised forensics, again, each element of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} of the target of the investigation D_{IT} and the investigation result D_{IR} and an element $f_{DF,\,SF}$ of the set of transfer functions $F_{DF,\,SF}$ need to be observed. This leads to the following definition:

Error occurs if for a given element the amount of the case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} in the data of the investigation result D_{IR} is larger than the data of the investigation target D_{IT} as a result of the inverse of a transfer function $f_{DF,\,SF}$ returning no result or a result outside of the set of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC}.

This is equivalent to the following formalisation (equation 4.10):

$$\forall\, d_{IT} \in D_{IR} \wedge d_{IT} \notin D_I \setminus D_{IFC} \wedge \forall d_{IT} \notin D_I \setminus D_{IFC} : \exists\, d_{IR} \notin D_{IT} \quad (4.10)$$

resulting in equation 4.11:

$$\exists\, d_{IR} \in D_{IT} \wedge d_{IR} \notin D_I \setminus D_{IFC} : f^{-1}_{DF,SF}(d_{IR}) = \{\} \quad (4.11)$$

In the case of error, also data of an investigation result d_{IR} can occur, which is outside the set of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC}, forming outlier data D_O (equation 4.12):

$$\exists\, d_{IR} \in D_{IT} \wedge d_{IR} \in D_I \setminus D_{IFC} \wedge d_o \in D_I \setminus D_{IFC} : f^{-1}_{DF,SF}(d_{IR}) = d_o \quad (4.12)$$

The resulting Venn diagram is depicted in the following Figure 23, where we define D_{IT} as a proper subset to D_{IT} and thank Sabah Jassim for this input of that nature during the Ph.D. defence.

.

[1] http://www.wireshark.org/

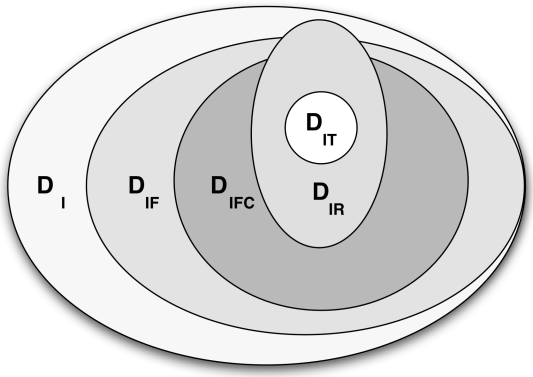

Figure 23: Relationship of the data of the investigation target D_{IT} and the investigation result D_{IR} in the case of *error*

Error is the violation of the security aspects of integrity and authenticity of forensically relevant data contained in IT systems caused by unintentional or purposeful false actions of hard- and software of an IT system or IT system operators, examiners and malicious intruders caused by vulnerabilities of design, implementation, configuration and social engineering according to the incident taxonomy [HoL12, pp. 6-17] enhanced in [KLD08, p. 414] outlined in Section 2.5.1.

In digital forensics error could for example result from a Memory Management Unit of the IT system, erroneously programmed by its operating system, when analysing a Main Memory Dump (see Section 4.2.1.2). With that, the memory dump under analysis as case-specific, forensically relevant information contained in case-specific, forensically relevant data is contaminated with data containing information from the IT system used for analysis.

An example for error in digitised forensics is when due to a faulty program logic the feature extraction used to extract fingerprint features when examining a latent fingerprint contained in acquisition data from a contact-less 3D surface measurement device (see for instance [HKG+11, p. 5]) is extracting the wrong features, resulting in a miss-classification between fingerprint residue and surface areas (see also Section 7.2). In this case, the wrong case-specific, forensically relevant information contained in case-specific, forensically relevant data as investigation results are returned from the transfer function, which do not have an associated investigation target as input data.

4.1.3 Uncertainty in digital and digitised forensics

To define uncertainty in digital and digitised forensics, again, each element of case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} of the data of the investigation target D_{IT} and the data of the investigation result D_{IR} and an element $f_{DF,\,SF}$ of the set of transfer functions $F_{DF,\,SF}$ need to be observed. We assume multiple runs of those transfer functions with the same data acting as data of the investigation target D_{IT} (also outlined in Section 4.4.1.1 as tool testing for digital forensics and in Section 3.4.1 with regards to ACE-V in digitised forensics). Of particular relevance is the relationship between two subsets of the data of the investigation result D_{IR}. This leads to the following definition:

Uncertainty occurs if for a given element the amount of the case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} in the data of the investigation target D_{IT} an element $f_{DF,\,SF}$ of the set of transfer functions $F_{DF,\,SF}$ yields two or more different elements of the case-specific, forensically relevant data containing case-specific, forensically relevant information D_{IFC} in the data of the investigation result D_{IR}.

This is equivalent to the following formalisation (equation 4.13):

$$D_{IR_1} \subset D_{IR} \wedge D_{IR_2} \subset D_{IR} \wedge D_{IR_1} \cap D_{IR_2} = \emptyset \wedge (d_{IR_1}, d_{IR_2}) \in D_{IR} : \\ \exists\, d_{IT} \in D_{IT} \wedge f_{DF,DF}(d_{IT}) = d_{IR_1} \wedge f_{DF,DF}(d_{IT}) = d_{IR_2} \qquad (4.13)$$

The resulting Venn diagram is depicted in the following Figure 24.

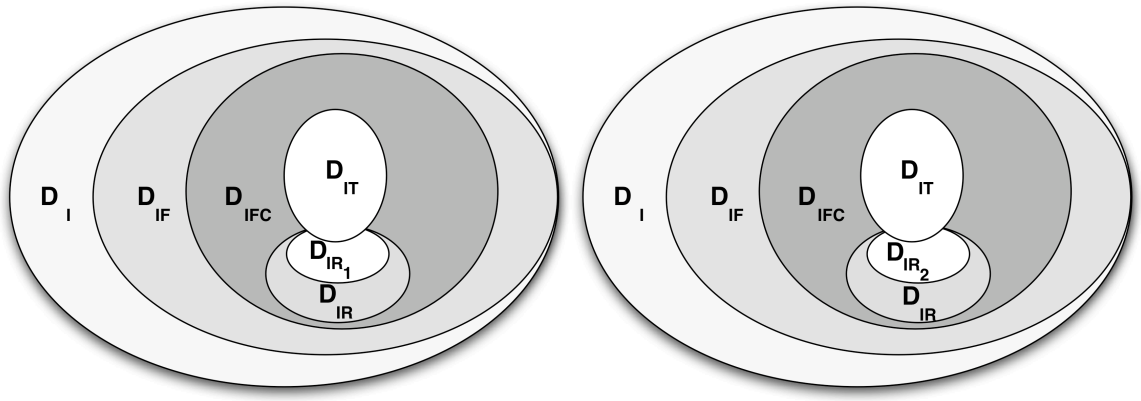

Figure 24: Relationship of the data of investigation target D_{IT} and data from the investigation result D_{IR} in the case of *uncertainty*, D_{IT} yields D_{IR_1} or D_{IR_2}

Uncertainty affects accuracy and integrity/consistency according to the Cyber Forensics Assurance model [Dar10, p. 64] (see Section 3.1). Its cause can be both normal and abnormal system use (unintentionally and intentionally) rooted in vulnerabilities of design, implementation, configuration and social engineering according to the incident taxonomy [HoL12, pp. 6-17] enhanced in [KLD08, p. 414] outlined in Section 2.5.1.

In digital forensics uncertainty for example could result from using flash-based solid state mass storage devices (SSD) as investigation targets to acquire case-specific, forensically relevant information contained in case-specific, forensically relevant data by creating a bitwise image. Simply powering on those devices could result in the alteration of the content by the firmware inside those devices [BeB10, p. 11]. A repeated acquisition with the same data of the investigation target D_{IT} would result in different case-specific, forensically relevant data D_{IFC} as data of the investigation result D_{IT}.

Since digitised forensics by its very nature relies on digital systems for its operations, it inherits uncertainties arising from the execution environment in both hard and software (e.g. time-delayed garbage collectors, memory errors etc.).

For the research presented in this thesis a top-down approach is used based on the data created, processed and communicated in an IT system. Other approaches, such as e.g. [Car06, p. 1-153] or [Gla04, pp. 75-118], employ a bottom-up view by looking at each state transition inside an IT system and inferring a history, thus realising an (semi-) autonomous event reconstruction. Such an approach is likely to be much more precise and works predominantly on the computer scientist's view (see [PBM08, p. 114]) in digital forensics. The shortcomings of such an approach include the problem of the complexity of real world systems and the computational complexity of the event reconstruction algorithm, which reaches from polynomial to exponential complexity, depending on the problem statement (see [Gla04, p. 162-163]). Also to establish formally e.g. if a file was created by a user with a web browser or because someone broke into the computer, is typically not possible following the hypothesis-based approach as presented by [Car06, p. 146].

Generally, DCEA can only address the balance between the known known and the known unknown types of knowledge (or lack of) as modified from [For14, p. 6-8] from a presentation given by former Secretary of Defence Donald Rumsfeld as follows:

- There are things that we *know that we know*,
- There are *known unknowns* - things we know that we do not know them,
- There are also *unknown unknowns* - things we do not know that we do not know.

The goal, next to a formal description of the ratio between the known unknowns and the known knowns, should be a shifting of balance to the latter.

4.1.4 Reflection and parting thoughts

With the formal definition of loss, error and uncertainty based on the data containing information, we established contribution C1, namely:

> C1:
> Formal description of loss, error and uncertainty using distinct sets of information contained in data (distinction into D_I, D_{IFC} D_{IFC} with only the latter solving specific incidents)

In order to establish a measure for loss, error and uncertainty in IT systems used for digital and digitised forensics as defined in the Sections 4.1.1-4.1.3, the following topics need to be researched:

- Forensic data types (representing the computer scientist's view [PBM08, p114]),
- Sets of methods (representing the computer scientist's view [PBM08, p114]),
- Sets of examination steps (representing the practitioner's view [PBM08, p114]).

Therefore, in the following sections the elements constituting the Data-Centric Examination Approach (DCEA) based on [KHD09, pp. 1-3] are introduced, which are designed to address the challenges centred on the concepts of loss, error and uncertainty as outlined in Section 4.1 on the conceptual level. Selected ideas used in the existing models (see Section 3.1) are modified and integrated into the approach.

Firstly, we look more detailed and practical into the structure of the data that forms the data of the investigation target D_{IT} and the data forming the investigation result D_{IR} as forensic data types.

4.2 Forensic data types

Forensic data types can help to categorise and chose the right tools (for all examination steps). The general idea is that data from the same data type are created, processed, stored and used similarly by the IT system in question and thus can be forensically gathered, investigated, analysed and documented in a similar manner. Even for unknown challenges, e.g. new, non-standard systems this categorisation can help to specify methods of the forensic process for their acquisition, investigation, analysis and documentation as well as strategic and operational preparation.

Establishing forensic data types can be seen as a part of the preservation of the data/tool sovereignty of the forensic examiner, serving a vital part of the research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

We extend the idea of forensic data types based on the initial idea communicated in [KHD09, pp. 2-3] and put them in the context of the formalisation of loss, error and uncertainty in a data-centric view of the forensic examination process as discussed in the preceding Section 4.1.

Formally we assign DT to the data types for digital forensics DF and DD to the data types of digitised forensics SF. We define for digital forensics DF the following set relation based on the set definitions from Section 4.1 (equation 4.14):

$$D_{IT}, D_{IR} \in DT \quad (4.14)$$

and for digitised forensics the set relation (equation 4.15):

$$D_{IT}, D_{IR} \in DD \quad (4.15)$$

to assure, that only those forensic data types form the input and output data space for operations in digital forensics DF and digitised forensics SF, respectively.

Following those findings, we define forensic data types as:

> Forensic data types:
>
> Forensic data types DT for digital forensics DF and forensic data types DD for digitised forensics SF represent the data in the investigation target D_{IT} and the data of the investigation result D_{IR}. They are ordered in layers that are not mutual exclusive; information contained in one layer could also be contained in a different layer. Information with the same semantics can be encoded in a different kind of syntax. Forensic data types are part of the set of data containing information D_I and part of the set of data containing forensically relevant information D_{IF} and most importantly for a successful examination step are also part of the set of case-specific, forensically relevant data D_{IFC}.

By using this data-centric approach based in the information contained in an IT system, we address demands for an information-centric asking for "rich" information that may be sufficient for the examination as demanded by [Pol08, pp. 20-21] (see also Section 3.1.5).

Our definition of forensic data types is based on the notion of the layer-based view from [Car03, pp. 3-7]. We use part M2 of our methodology (see Section 1.2):

> M2:
>
> We apply the selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 621] to data stored, processed, communicated in IT systems to distinguish forensic data types for digital and digitised forensics. To achieve this, we construct layers of data by giving them semantics that support the forensic process. In further conjunction with the ISO/OSI model, we use a layering that is not mutual exclusive.

We use the following characteristics embodied in the ISO/OSI model of the ISO/IEC 7498 ([Cas11, p. 621] based on [ISO19]):

- layering of data by giving them semantics with regards to a specific purpose,
- layering of data in a way that the same semantics can be present in different syntactical encapsulations.

Incorporating semantics can support addressing item IIIa Availability/Timeliness ("whether they were true for the given timeframe and environment") and item IIIb Utility/Relevance ("Is it useful / is it the right information?") of the cyber forensic assurance [Dar10, pp. 62-63]. In addressing semantics, we follow the demand that the meaning of data needs to be understood [Cas02, p. 32]. The forensic data types from the basis for intelligent, analytical approaches reaching further than:

- literal string searching,
- simple pattern matching,
- indexing data to speed up searching and matching,
- hash analyses,
- logical file reviews

as suggested by [Bee09, p. 26]. In the following, we incorporate the identified characteristics of the ISO/OSI model of the ISO/IEC 7498 ([Cas11, p. 621] based on [ISO19]) in a layering of data containing information D_I that serves forensic purposes and can be used to describe both the data of the investigation target D_{IT} (source) and the data of the investigation result D_{IR} (outcome).

The data types are application/use case specific. The data types described in this thesis are shown to fit the exemplary chosen use cases for digital forensics. Other targets (e.g. Industrial Control Systems) may need an addition of data types or alternative definitions (e.g. [AHK+19, p. 135]).

Digitised forensics needs entirely new data types (see Section 4.2.2), however, we apply the same characteristics from the ISO/OSI model of the ISO/IEC 7498 ([Cas11, p. 621] based on [ISO19]).

4.2.1 Forensic data types for digital forensics

We use our existing findings from [KHD09] and our publications based on them [KHD+09], [KHA+10], [HKD11], [BKS13], [KDV15], [AHK+19] and extend them to the data stream idea, where all of the forensic data types form a common forensic data source and check their applicability in this new context. Also we check whether the definitions need updating. In this section we use M2 as our methodology (Section 1.2) and achieve contribution C2.1 (Section 1.3).

In [KHD09, pp. 2-3] 8 data types are identified as being forensically relevant. For this thesis we maintain those data types and their description. However, we change the ordering of two data types based on the fact that all information can be extracted from the raw data, giving raw data a special significance as the root of all data types thereafter (Table 3).

Forensic data type	**Description according to [KHD09]**
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified
Hardware data (DT_2)	Data not or only in a limited way influenced by the OS and application
Details about data (DT_3)	Meta data describing other data
Configuration data (DT_4)	Modify the behaviour of the system and applications
Communication protocol data (DT_5)	Modify the communication behaviour of the system
Process data (DT_6)	Data about a running process
Session data (DT_7)	Data collected by a system during a session
User data (DT_8)	Contents created, edited or consumed by the user

Table 3: Forensic data types $DT_{1\text{-}8}$ as defined in [KHD09, pp. 2-3] with altered ordering, given the significance of raw data as the source of all data types

Based on the nature of data provision (time-discrete, i.e. the actual information position is known in full at any given time, e.g. can be represented as a matrix vs. time-continuous, i.e. the actual information is only partially known as bits travelling a linear communication channel, see also [Hil20]) and the data volatility, we suggest a distinction into three different data streams DS:

- Mass storage data stream DS_T: Time-discrete source of data from a computer system stored and manipulated internally and externally, typically with a lower volatility, usually long term data retention
- Main memory data stream DS_M: Time-discrete source of data from a computer system, stored and manipulated internally, source of typically highly volatile data, usually short time data retention
- Network data stream DS_N: Time-continuous source of data from a computer system communicated externally, highly volatile, short time data retention

As shown in Figure 25, the forensic data source includes data acquired from all data streams, although sometimes only one stream is used. However the correlation and/or fusing of data streams from different streams is likely to yield better results because of better completeness based on forensic source (for both inculpatory and exculpatory evidence, see Section 2.7), as shown in our work in [KHA+09, pp. 480-487] where a correlation from both the main memory data stream DS_M and the mass storage data stream DS_T yields portions of data normally lost due to the specific deletion mechanism of the ext3 filesystem (see also Section 4.2.1.1.2).

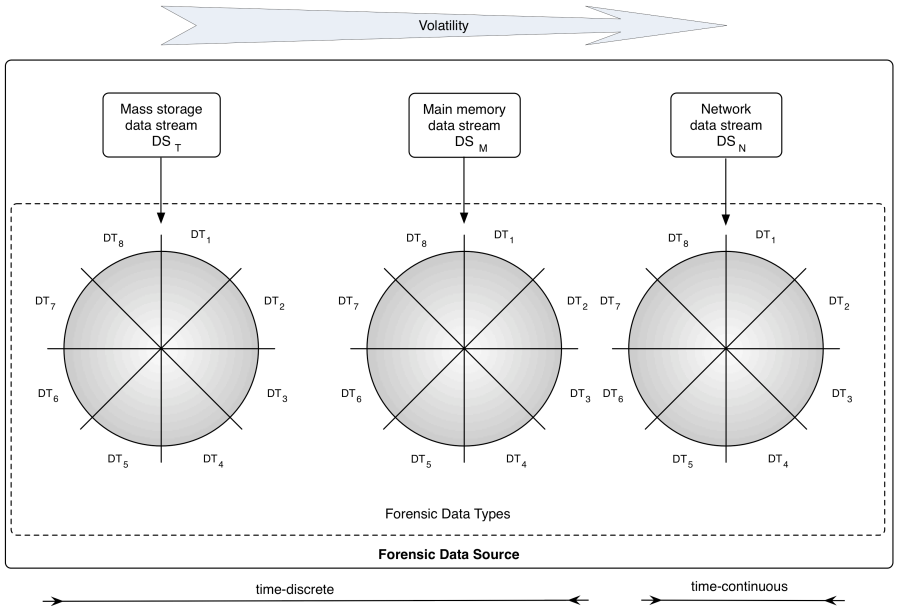

Figure 25: Forensic Data Source separated into streams for digital forensics DF and according to time-discrete and time-continuous characteristics

In the following we will look for each stream into the selected aspects of the general structure of the data organisation, useful characteristics for digital forensics and provide examples of forensic data types.

The forensic data types and their data source are, as already pointed out in Section 4.2 are different for the application/use case of digitised forensics. In the following we introduce data streams for digital forensics and discuss data types for digital forensics based on [KDV+15], which summarises our state of the art prior to this thesis. In due course we propose a few minor alterations and direct the user to application areas, where an extra data type is required [AHK+19, pp. 128-136].

4.2.1.1 Selected fundamentals for mass storage device forensics

Mass storage device forensics is probably the oldest discipline targeting the mass storage data stream that is one of lesser volatility variety, if further write access to the given device is prevented. To achieve this, software-based solutions (often specially modified bootable operating environments (e.g. [Bas19]) and hardware-based devices are available (see e.g. [BuW06, p. 129]). The hardware-based solutions are to be preferred in general since they can filter out commands that constitute a write access at device interface level and are less error-prone (e.g. if the bootup-sequence of the suspect IT system in BIOS/EFI is selected erroneously). However, in specific situations (see Section 6), even the usage of such a blocking mechanism cannot prevent content alterations.
To gain an understanding regarding mass storage device forensics, in the following Section 4.2.1.1.1 the general structure is explained, with which we will look for useful characteristics in Section 4.2.1.1.2 context with the forensic data types outlined in Table 3.

4.2.1.1.1 Generalised structure of data organisation in mass storage devices

The general concept of mass storage device is that of a *volume*. According to [Car05, p. 57], a volume is a set of addressable sectors that are accessed by the operating system or in some cases by the applications themselves. This general concept is not bound to any particular physical storage and read/write mechanism (e.g. electro-magnetic storage on platters or tapes, non-volatile storage on isolated charges in memory cells, changes in reflective capacity). Different access types are available. For magnetic discs as used in hard disk drives or Flash memory/EEPROM based devices such as Solid State Drives (SSD) or USB thumb drives, random access strategies

are available. However, in media primarily used for archival purposes, e.g. backup tape, (re-) recordable CD/DVD, sequential access strategies are employed, either for recording or both recording and playback alike (see also the appendix Sections 10.4.1 and 10.4.2).

Volumes form a tree-like structure with the physical device forming the *parent volume* (see Figure 26). For a number of reasons (e.g. storage space management, access control, provision of swap space etc.), this parent volume is often partitioned into a number of volumes located in the parent volume. This instance of a volume is sometimes also called a *partition*.

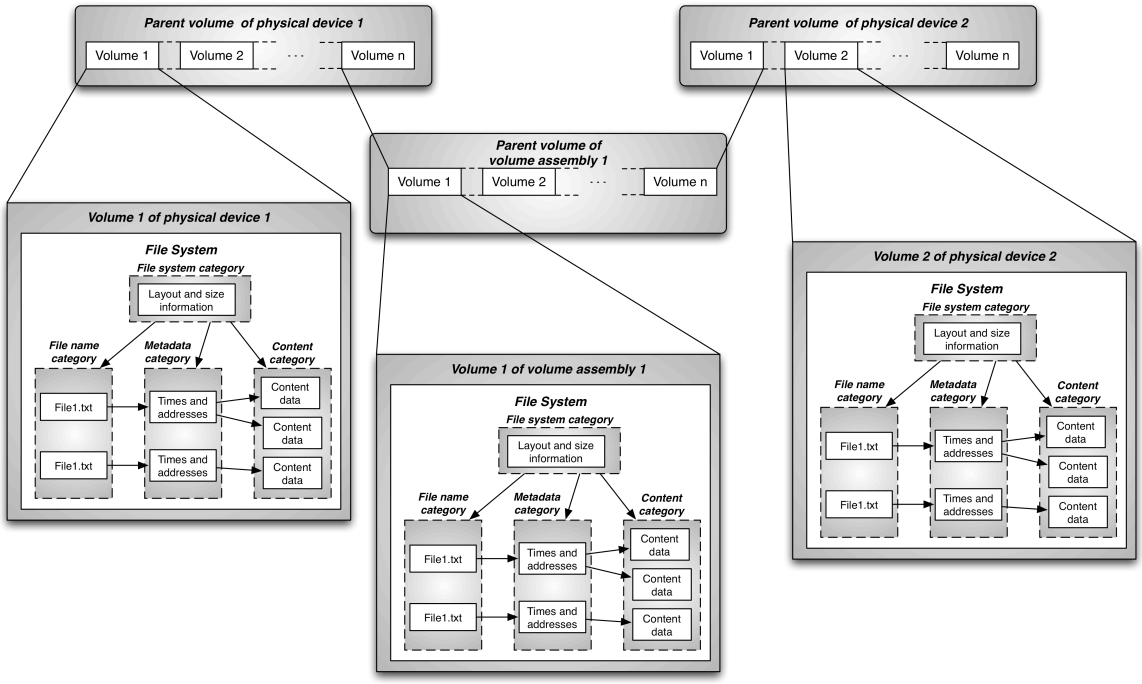

Figure 26: Simplified abstract view on mass storage data organisation modified from [Car05, p.58] with the volume organisation as a tree structure starting from a parent volume and with the consideration of volume assemblies [Car05, p. 60], i.e. more than one physical device forming a parent volume

All those child volumes are managed in the *partition table* with its main important entries being its starting and ending sector. This table is typically stored in the starting sectors of the parent volume. These are the only trustworthy entries according to [Car05, p. 58] on the grounds that the IT system needs them to work. All other values (e.g. partition type) can be fake on purpose or attributed falsely by oversight.

For example, on IA32 architectures with DOS disklabel (partition type), the Solaris operating system used the partition type number 82 as its identifier [Ora19, p. 121]. This number is also assigned to a Linux swap partition. So even booting a live Linux system (e.g. from removable media such as DVD-ROM or USB thumb drive) could lead to the destruction of a Solaris disk structure since the existence of a type 82 partition often initiates an initialisation of that partition as swap space.

Logically, the *volume* is a set of consecutive sectors. On the physical device, the sectors of a parent volume do not necessarily have to be consecutive. In fact, as shown in Section 6 a self-similar structure can be found in the physical device itself, that projects a view of a consistent number of sectors whilst maintaining spare sectors, service areas etc. whilst internally constituting a processor-controlled component.

A technique called *volume assembly* allows multiple physical devices to be treated as one logical parent volume, to which all of the aforementioned techniques can be applied. Typical reasons are added redundancy through managed parallel operation (in case a drive fails) or added capacity in

managed serial operation (the capacity is the added capacity of the individual physical devices). Also, a combination of those to approaches is possible and in use. A very popular example of logically tying physical drives together is the *RAID* mechanism, which stands for *Redundant Array of Inexpensive Disks* (see [Car05, p. 111]), which can be implemented with added hardware or in software alone.

Obviously, the configuration of the volume structure constitutes very important configuration data, which is relevant for the forensic data types of the mass storage data stream introduced in Section 4.2.1.

File systems inside of volumes provide a mechanism for the operating system and the application to store data in a hierarchy of files and directories as a means for long-term data storage and retrieval. To achieve this goal, file systems consist of structural and user data that are organised for easy access by the IT system [Car05, p. 129].

According to [Car05, p. 130], file systems can be structured into categories according to their role in the file system as follows (see also Figure Figure 26):

- File system category - contains general file system information, essentially a map of the file system, forensically relevant information include the size of data units and where to find certain data structures,
- Content category - contains data that forms the actual file content, typically organised in a number of containers of fixed size as data units,
- Metadata category - contains data that describes a file or directory, including data unit location, access time, permissions etc. but not necessarily the file name,
- File name category - represents the human interface to the file system and is basically a lookup table to add a human readable file name that points internally to the meta data and thus in turn to the data units,
- Application category - this category represents help and assistance functions that are advantageous to have but are not required for the essential function of the file system (i.e. load and save files), examples are quotas or journals.

The application category is the only category regarding file systems that can be attributed as being non-essential (see Section 2.7.2) and thus having a considerably less evidence value compared to all the other categories.

In the appendix Section 10.4.1 we extend our generalisation to include systems used for backup/archival purposes.

4.2.1.1.2 Useful characteristics of mass storage management for forensics and mapping to forensic data types

Generally, forensics in mass storage exploits the fact that most data is never really deleted but rather marked invalid in the *file system category* (see Section 4.2.1.1.1) and left up to be overwritten if that bit of storage space is needed. This characteristic is used to speed up storage processes. Depending on usage characteristics, deleted elements (e.g. file data, partition data) can remain untouched for quite some time, ranging from hours to years, as detailed studies e.g. in [FaV05, pp. 153-154] have shown. Thus, largely intact files can potentially be recovered. They can cover all forensic data types from Table 3 in Section 4.2.1 that typically have a file system representation (see Table 4), often ruling out process data DT_6 and hardware data DT_2.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	File system representation as swap file/hibernation file
Details about data (DT_3)	Meta data describing other data	File system table meta properties as file system representation, e.g. $MFT, $Volume for NTFS
Configuration data (DT_4)	Modify the behaviour of the system and applications	File system representation for preference data e.g. system preferences data
Communication protocol data (DT_5)	Modify the communication behaviour of the system	File system representation for network communication configuration e.g. network settings file
Session data (DT_7)	Data collected by a system during a session	File system representation of log files especially covering user and network sessions
User data (DT_8)	Contents created, edited or consumed by the user	All user content represented as files in a filesystem

Table 4: Forensic data types as defined in [KHD09, pp. 2-3], which are likely to have a file system representation

Further, areas on a volume exist, which constitute residual data (ambient data) and are not actively used by an IT system [New07, p. 252]. These are known as *slack spaces*. Slack spaces typically occur between *partitions* (volumes, see Section 4.2.1.1.1) or within a file system in the *content category* (see Section 4.2.1.1.1) on a given parent volume [Lew04, p. 5]. Volume slack occurs when a parent volume is repartitioned. According to [Lew04, p. 5], the unoccupied space between volumes can contain remnants of old or deleted files and other file system artefacts. File slack occurs due to the grouping of physical sectors (device specific but often 512 bytes) into clusters (file system specific, often 4096 bytes). Those clusters then form the smallest addressable unit. As a result, typically a difference between the logical and the physical end of a file exists (e.g. only using 42 bytes off the 4096 bytes of the last cluster of a given file), which contains the previous data of said cluster [Lew04, p. 5]. Since there is typically no automatic interpretation of those portions of previous data (i.e. slack spaces), we define them as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 5 shows.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Unstructured partial residual data that previously belonged to files in the filesystem

Table 5: Slack spaces as raw data DT_1 as defined in [KHD09, pp. 2-3]

A lot of modern file systems also employ the technique of maintaining a *file system journal*. Belonging to the *application category* (see Section 4.2.1.1.1), a journal is essentially a log of intentions regarding the meta data of the file system (see [Car05, p. 152]). Any changes to those meta data (e.g. during deletion, append or generally write actions) are recorded prior to their execution. In the event of e.g. a crash or power failure, the subsequent file system check is faster since the log of unfinished transactions on the file system is consulted, resulting in a decision either to complete the action if sufficient data is already written to the content category or the whole alteration from before the crash is reversed. Some of the journal content can potentially be recovered (regardless of the presence of a crash), thus providing information of previous states of the file

system. However, it needs to be stated, that journals form non-essential data (see Section 2.7.2) and thus can be modified or deleted by attackers without interfering with the normal function of the file system. The defining characteristic of this log data containing meta file information is, that it is maintained in a session, thus we define this data as session data according to the data type definition in Table 3 in Section 4.2.1, as the following Table 6 shows.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Session data (DT_7)	Data collected by a system during a session	Journal logs of modifying actions on filesystems, e.g. deletion, addition to existing files, creation of new files

Table 6: File system journal logs as session data DT_7 as defined in [KHD09, pp. 2-3]

Another characteristic that can be of great importance for forensics from the *content category* (see Section 4.2.1.1.1) is the fact, that most existing file systems typically place the content sections of a file in a contiguous order during a write process, if the file system is not heavily fragmented. The latter occurs typically if the file system is close to its capacity limits and a free set of blocks needs to be identified anywhere in the unallocated sector area known to the file system. This characteristic of contiguous ordered sectors used to store content is exploited by the forensic technique of *file carving* [Ham18, p. 188], often used as a last resort mechanism to recover some sort of structure (an thus semantics) from a mass storage device, whose file system structure is either substantially corrupted or if the file system type is of an unknown type. Files often have a standard header (or signature) and sometimes footer. File carving looks for a given header and footer (a type-specific fixed sequence of bytes) and copies the content between header and footer into a file [Ham18, p. 188]. Lacking a footer for a given, a fixed size of bytes after the header detection can be submitted to the tool instead. The great disadvantage of file carving is the high number of false positives. Any occurrence of the header is treated as the beginning of a file although those sequences could easily represent other semantics (e.g. part of an audio file sample content). Often the ratio of false positives is higher by magnitudes compared to true positives. Another disadvantage is the lack of any meta data when solely relying on file carving. By its very nature, file carving operates on data that we define raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 7 shows, since there is typically no automatic interpretation of the data source file carving operates on, otherwise more effective data recovery options could be used instead.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	File carving operating on unstructured data areas to gain some structure based on searchable characteristics (e.g. header, footer)

Table 7: Data source for file caring as raw data DT_1 as defined in [KHD09, pp. 2-3]

The virtual memory system (see also Section 4.2.1.2.1) employed by the majority of today's IT systems also provides a *swapping mechanism*, which is a characteristic exploitable by forensics. It can operate on the *content category* or the *parent volume* (see Section 4.2.1.1.1) of the generalised structure of data organisation in mass storage devices. This enables the virtual memory manager to store main memory pages onto a dedicated area on mass storage media if certain conditions are met (FaV05, 164). This mechanism is used e.g. when the virtual memory manager runs out of free main memory pages. Depending on the operating system and its settings, the swap area can be a regular file in the file system or a dedicated swap partition. Those swapped out memory pages can contain all sorts of potential case relevant data (e.g. keys used for encryption, pass-

words etc.). Also, when memory pages are swapped out to mass storage media, they have a lesser degree of volatility (see Section 3.3.1.1), improving the chances of their forensic recovery. A somewhat similar mechanism to swapping is *hibernation* [BuW06, p. 381], which is typically used in mobile devices (e.g. laptops). Here, a system can resume operations quickly after waking up from being suspended beforehand. If a system is placed into hibernation mode, the whole content of the main memory is placed onto mass storage (typically as a regular file in the file system). It has a size equal to the amount of main memory. Similar to the swap mechanism, this file contains main memory organised in pages and thus similar types of data can potentially be retrieved. Gathering data from swapping and hibernation mechanisms provides data that we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 8 shows.

Forensic data type	**Description according to [KHD09]**	**Context in mass storage forensics characteristics**
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Swap and hibernation files need context from the memory manager of the operating system to have inherent intelligible structure, this is typically unavailable in post mortem forensics

Table 8: Swapping and hibernation mechanisms as a source for raw data DT_1 as defined in [KHD09, pp. 2-3]

Multiple mass storage devices can be treated as one *logical parent volume* (see Section 4.2.1.1.1), forming a volume assembly. Since many of the different types of volume assembly (e.g. a RAID formation) require physical volumes with an identical amount of storage capacity, mechanisms to (typically) reduce the storage space reported from a given device exist. Such a mechanism is, for instance, the *Drive Configuration Overlay DCO* (see [Car05, p. 37], which is standardised in the ATA-6 command set for mass storage devices and allows (amongst others) for the reduction of physical sectors and thus to mask out (i.e. hide) portions of the physically available sectors of the storage number. Another option, even earlier added in the ATA-4 command set, is the establishment of a *Host Protected Area HPA* (see [Car05, p. 36]). Designed to mask out sectors using the standard means of access, it allows for the full formatting of the sectors left to operating system access whilst maintaining e.g. restore data when accessed by a BIOS routine. Both HPA and DCO have legitimate uses as described here but can also be used for malicious uses. All of the mechanisms mentioned in this paragraph we define as configuration data DT_4, according to the data type definition in Table 3 in Section 4.2.1, as the following Table 9 shows.

Forensic data type	**Description according to [KHD09]**	**Context in mass storage forensics characteristics**
Configuration data (DT_4)	Modify the behaviour of the system and applications	Mass storage media have the option of configuring available disk storage space

Table 9: Mass storage device size alteration techniques using configuration data DT_4 as defined in [KHD09, pp. 2-3]

File systems, in order to function or as added support for applications and the user, add and maintain data as described by the *file name* and *meta data category* (see Section 4.2.1.1.1). Of great importance is meta data that is *essential* to the functioning of the file system as it is harder to manipulate and maintain a working system (see Section 2.7.2). These include e.g. file names assigned to the portions from the content category that make up said file. Without the correct links, the file system would behave erroneously. *Non-essential data* can still be of importance for the event reconstruction. Great care however needs to be taken that no signs of tampering with that data took place before drawing conclusions. These include *permissions* to file system objects such as files and folder. For each of those objects, although often file system specific additions are

implemented, on the most basic level, the permission of read, write and execute are maintained (see e.g. [FaV05, p. 49]). Also, the *ownership* of an object of a file system can yield critically information. Typically numerically coded, it links a file to a given username of a system and a group respectively [FaV05, p. 48]. Especially the privilege escalation towards super user state or addition to a group that has been granted access to privileged resources is often parts of a successful system intrusion. Especially important are Modify/Access/Change MAC Times for forensic timelining (see also Section 3.3.1.4). Although, again, being file system specific, generally it can be stated (see [FaV05, p. 18] that the Modify time refers to the last event, which changed a file or folder. The Access time refers to the last use of that object while the Change time (in some file system referred to as Create time) describes the last changes in the meta data (e.g. change of ownership, permission etc.) and consequently, the creation of a file results in said changes. Meta data are an important source of information regarding events in a system. We define them as details about data DT_3, according to the data type definition in Table 3 in Section 4.2.1, as the following Table 10 shows.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Details about data (DT_3)	Meta data describing other data	File system maintains extra (non-functional) data describing properties of files

Table 10: File system meta data as details about data DT_3 as defined in [KHD09, pp. 2-3]

All of the above can also be used to describe mass storage devices often used in virtualised environments (e.g. emulators, virtual machines etc.) with the only difference that the consecutive number of addressable sectors is provided by a regular file within the host file system that is presented by the virtualisation/emulation software as a block device to the virtualised/emulated guest environment.

In media primarily used for archival purposes, e.g. backup tapes, (re-) recordable CDs/DVDs, sequential access strategies are employed, either for recording or both recording and playback alike, whose properties can also be described using the proposed forensic data types (see appendix Section 10.4.2).

In the following we look at the next data stream of main memory DS_M for both fundamentals and useful characteristics for digital forensics.

4.2.1.2 Selected fundamentals of main memory forensics

Main memory forensics is one of the younger branches of digital forensics. Examining the memory stream of an IT system can be important e.g. to circumvent encryption and to detect malicious software that leaves no traces on non-volatile memory (e.g. mass-storage devices).

It also contains details such as running processes, open files, loaded libraries, network sockets etc. and can deliver even more information when used in conjunction with mass storage forensics (Section 4.2.1.1) and network forensics (Section 4.2.1.3).

However, main memory forensics is notoriously difficult to execute. Contrary to mass storage forensics, where data is stored intentionally structured and often well documented to be read back (potentially by other applications), the data stored in memory is solely used by the particular running process. Hence, it is typically using *undocumented data structures*, which may also frequently change during different versions of an application [Gar10, pp. S67].

Main memory content typically constitutes *highly volatile data* (see also Section 3.3.1.1). Firstly, all content is lost when the power supply to the IT system is interrupted (by an orderly shutdown or "pulling the plug"). Also, during normal operations IT system the content of the main memory stream changes a lot more frequently than the content of the mass storage stream. A number of entities (e.g. operating system, IT applications, hardware devices using Direct Memory Access)

can change the state of main memory in a second. To understand the underlying problems of main memory forensic analysis, one has to take a look at how main memory is typically organised in IT systems. For that, in the following the abstract (system) view of physical address space and the process view of virtual memory are illustrated (see also Section 5). Memory organisation and the data contained therein is also highly dependent on the operating system and the underlying hardware, which is why a) we can only give a coarse, generalised view on the system and can only pick selected, representative facts in the following subchapters. The reader is instructed to consult system-specific literature (e.g. [FaV05, pp. 87-116], [LCL+14], pp. 117-858).

4.2.1.2.1 Generalised structure of data organisation in main memory devices

Modern CPUs employ a special unit called the memory management unit (MMU). Its purpose is to provide and support translations from virtual to physical address spaces (see also [Tan01, pp. 202-205]). The need for that *memory management mechanism* arises from the multitasking property of modern operating systems. It provides each running process with a continuous addressable memory space according to its allocation demands by mapping a number pages of physical memory to into that address space [PWF+06, p. 198]. Those pages are fixed sized units (typically 4k or 8k in size). Processes can require more memory than is physically available, especially in a multitasking environment with a multitude of processes loaded at the same time. Therefore cheaper but slower virtual memory (typically implemented as a special file or partition on mass storage devices) is used to swap out areas of inactive regions of application memory. So a virtual address can refer to a physical memory cell or a swapped out area on mass storage devices (see Figure 27).

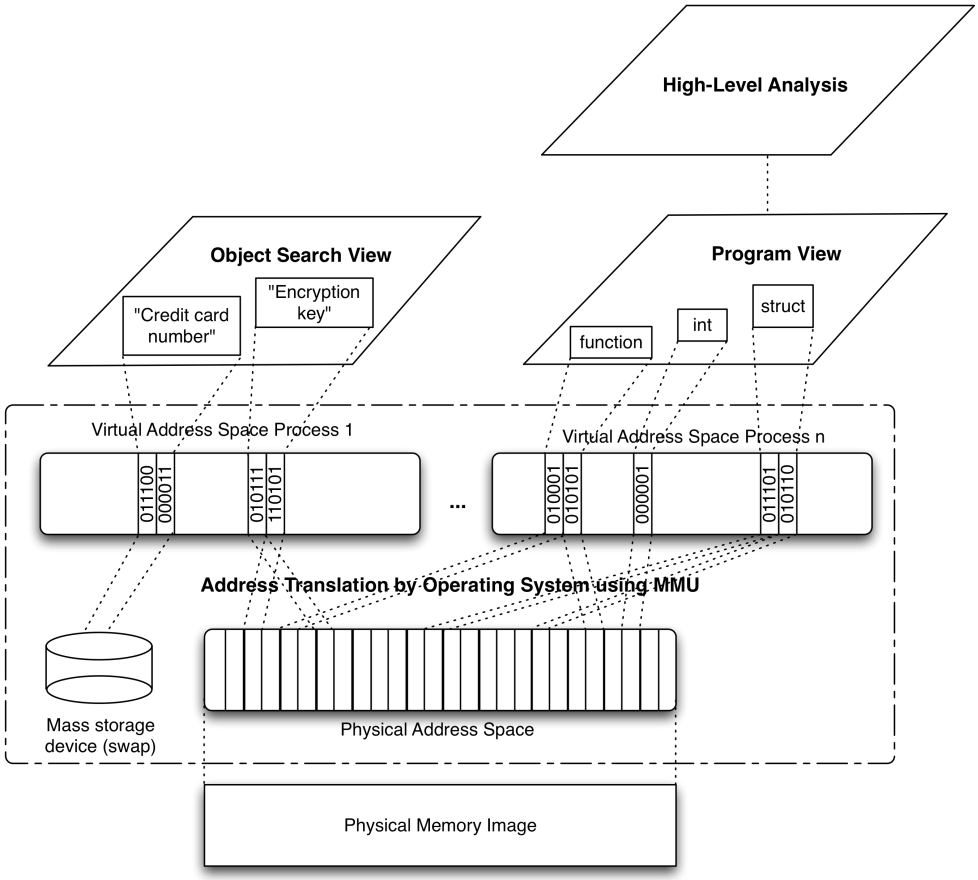

Figure 27: Connection of physical main memory image, address translation and forensic views (enhanced from [PWF+06, p. 201])

In the latter case an exception is triggered by the operating system to reload the content into memory by either copying the content into a page marked as unused when there is free space available or by initiating the swap out procedure on currently unused sections of other processes.

In order to obtain structure from the low-level data, according to [PWF+06, p. 199] it has to be identified, how software organises and interprets data within its own view of memory.

A *process*, as the most central concept in operating systems, is an abstraction of a running program, including the current values of the program counter, registers and variables [Tan01, p. 72]. The operating system itself is a program, organised in processes, too. By saving the contents of the - highly volatile - CPU registers such as the program counter and the other registers and recent contents of variables to the process before activating a different process, the actual state of the process that has been pre-empted can be restored. Each process therefore has its own, virtual CPU [Tan01, p. 72] and by using the physical-to-virtual address translation, each process also has its own virtual memory as well (see also Figure 27). In main memory forensics the goal is to reconstruct the virtual memory for the individual processes. How virtual process memory is organised, depends highly on the IT system and CPU architecture, the operating system (and therefore the execution header, e.g. the Portable Execution Header (PE-header) [TIS93, pp. 1-80], the Executable and Linkable Format (ELF) [TIS93a, pp. 1-82]), the compiler used to create the program and the data structures. The latter is determined by the programming language and program source code.

Threads are designed to allow multiple executions in the same process environment, sharing the same address space and global variables [Tan01, p. 81]. They are executed quasi-parallel (as processes themselves are) and to a large independent from one another. Threads allow sharing an address space and all of its resources (e.g. open files). In Figure 27, a threat is a code entity with its own program counter, its own stack and registers holding its current variables [Tan01, p. 81]. Threads can in theory overwrite the stack of other threads, local variables etc., making them a target for malware investigation as part of memory forensics.

4.2.1.2.2 Useful characteristics of main memory management for forensics and mapping to forensic data types

The whole approach of memory management is somewhat comparable to the management of free sectors on mass storage devices (see Section 4.2.1.1). Analogous to the handling of free sectors the content of deleted pages is not physically erased, instead the page is marked as available for next use (de-allocation). For as long as it is not requested and overwritten, the content of that page can remain persistent for quite some time, depending on the use of the IT system as has been illustrated in [FaV05, p. 177]. Forensically relevant data can be contained in both the pages in active use (about running processes) and about pages still not overwritten, belonging to old processes. Additionally, for the most comprehensive forensic view, not only the physical main memory image but also the swap file or partition needs to be analysed. From those pages, primarily data we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 11 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Dereferenced memory pages and memory dumps per se are raw data

Table 11: Memory pages and memory dumps as a source for raw data DT_1 as defined in [KHD09, pp. 2-3]

The main memory content can be extracted (dumped) using hardware and software solutions with differing impacts towards the integrity of the collected data (see Section 3.3.1.5). Raw data is often the base for all following examinations, where this data is then interpreted, processed and formatted to reveal potential case relevant information. The memory dump data also falls into the category of data, which we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as Table 11 shows.

In main memory hardware information is recorded e.g. in Linux from drivers and kernel components as part of the debug ring buffer [LCL+14, p. 663]. Such information can be used to enumerate hardware/firmware component present in the IT system as recognised by the operating system. Also the Windows Operating system family typically stores hardware information in the system registry [Red19, pp. 32-35], which is also maintained in memory [LCL+14, p. 281]. We will use this combination for our research described in Section 5. As stated in [LCL+14, p. 663] this information can also be used to establish the presence of removable devices (e.g. mass storage devices such as thumb drives, external hard disk drives, character input devices such as keyboard, mice, dongles etc.) and extra information about them (e.g. firmware reversion, serial number). However, as shown in Section 6, this information can only be regarded as being circumstantial (i.e. of limited evidentiary value) in the light of the finding that this information can be changed by a reasonably skilled adversary. We define the data as hardware data DT_2 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 12 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Hardware data (DT_2)	Data not or only in a limited way influenced by the OS and application	Enumeration of hardware data in main memory by the operating system

Table 12: Hardware data DT_2 enumerated in RAM as defined in [KHD09, pp. 2-3]

Processes have *extra information* attached to them. Again, similar to mass storage, such information can contain meta data, which can aid a forensic examination.

Typically a *Process ID (PID)* is created, and the *start time* and if applicable, the *exit time* is recorded, (see e.g. [LCL+14, pp. 613-615]). Exited processes are kept, in analogy to deleted files in filesystems (see Section 4.2.1.1.2). Further, the extra information can contain *user information* such as UserID and Group ID. Meta data are an important source of information regarding events in a system. We define those meta data as details about data DT_3 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 13 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Details about data (DT_3)	Meta data describing other data	Retention of exited processes, maintenance of starting and exit times and users / groups

Table 13: RAM process data as details about data DT_3 as defined in [KHD09, pp. 2-3]

Also, *arguments* passed to the process are recorded, which often in cases of malware consist of traces like suspicious folder names, configuration options, variables etc. (see also [LCL+14, pp. 613-614]. We define them as configuration data DT_4 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 14 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Configuration data (DT_4)	Modify the behaviour of the system and applications	Arguments, variables to configure aspects of a process (e.g. read/write locations)

Table 14: Arguments, parameters passed to processes in RAM as configuration data DT_4 as defined in [KHD09, pp. 2-3]

The network configuration of an IT system and its networking properties are present in the main memory stream. Memory forensics can greatly enhance examination results due to the fact that up-to-date network information is often present in main memory only [LCL+14, p. 637]. Further, since querying the memory directly, as opposed to using live forensic tools that rely on the operat-

ing systems application programming interfaces (API), this examination strategy bypasses potential anti-forensic techniques that modify said APIs. That networking information we define as communication protocol data DT_5 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 15 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Communication protocol data (DT_5)	Modify the communication behaviour of the system	Up to date network configuration and properties in RAM

Table 15: Networking properties assigned as communication protocol data DT_5 as defined in [KHD09, pp. 2-3]

Further, process build a *hierarchy*, in which a parent process can fork other processes (children), see e.g. [Tan01, p. 76]. Further, they are typically connected to a specific user account and other, potential case-specific information. These and other relationships can typically be recovered, enabling the detection of anomalies, such as hidden processes.

The maintenance of processes by the operating system creates data we define as process data DT_6 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 16 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Process data (DT_6)	Data about a running process	Process hierarchy, user information, permissions etc. managed by the operating system controlling the processes

Table 16: Process maintenance and control data assigned as process data DT_6 as defined in [KHD09, pp. 2-3]

In some operating systems, a complete user *session* can be derived from information contained in main memory. They can chain together processes from the same session and can store information on a user's logon session and graphical user interface (GUI) objects [LCL+14, p. 152-153]. Operating on an inter-process scale we define those data as session data DT_7 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 17 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
Session data (DT_7)	Data collected by a system during a session	Multitude of processes in the same user session context, e.g. logon management by the operating system

Table 17: Data about the processes linked together to form a user session as session data DT_7 as defined in [KHD09, pp. 2-3]

(Parts of) *User* data in active processing is kept in main memory as part of a process opening or editing it, as a list of open files and links to memory containing it kept [LCL+14, p. 493]. They can be linked to processes (and thus, in turn to user accounts and other information) and thus provide better evidentiary value (see Section 2.7.2) then some file content just residing somewhere in the main memory of the IT system. We exploit this fact for the scenario described in Section 5. We define any user content as user data DT_8 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 18 shows.

Forensic data type	Description according to [KHD09]	Context in main memory forensics characteristics
User data (DT_8)	Contents created, edited or consumed by the user	Each opened user content resides (at least partially) in main memory

Table 18: User content in main memory as user data DT_8 as defined in [KHD09, pp. 2-3]

Further, it is possible to maintain a swap space directly in main memory (see e.g. [RiC14, p. S6]) as opposed to swap space on mass storage devices (see Section 4.2.1.1.1), although typically on smaller scale, storage capacity wise. This technique is called compressed RAM [RiC14, p. S4] and is achieved by compressing pages not currently being referenced. It adds a new opportunities to forensics as this swap space (although as volatile in nature as main memory itself) is not encrypted and can correlate memory pages swapped out with processes, giving traces contained in those pages much more evidentiary value [RiC14, p. S11].

Overarching the data streams within an IT system, the main memory content also includes information about open files and data contained therein as well as existing network connections and the data transmitted, greatly substantiating findings from those streams (see Sections 4.2.1.1 and 4.2.1.3). In the following we look at the network data stream DS_N for both fundamentals and useful characteristics for digital forensics.

4.2.1.3 Selected fundamentals for network forensics

Network forensics constitutes a fairly well researched discipline in digital forensics. It is younger than mass storage forensics (see [Cas11, p. 607]) but older than main memory forensics. On its own, i.e. with regards to communication media such as copper wires, the airwaves or fibre optic cables, it is of a very, very high volatility (see also Section 3.3.1.1). However, since networking devices such as network interface cards (NIC, see e.g. [Cas11, p. 609]) often have their own, albeit small, memory in the shape of buffers and report directly to process structures in main memory or influence content of log files, preference files etc. they also affect mass storage memory of a number of involved devices.

In the following, we will look into the generalised structure of data organisation in network forensics. Primarily we will use standard desktop IT networking based on IP based communication. In the appendix in Section 10.4.3 will also add notes for other network systems e.g. field bus systems such as the Controller Area Network (CAN) that is widely used in automotive systems (see e.g. [MöH19, p. 114] or [HKD11a, p. 11]).

4.2.1.3.1 Generalised structure of data organisation in network forensics

Networks are a system in which a group of elements (nodes) can exchange information via a transportation medium [Bos07, p. 70]. As such, networks can be described according to their topology. A network topology according to is a structure consisting of network nodes and connections. A specific network topology is typically (see Figure 28) based on basic topology types.

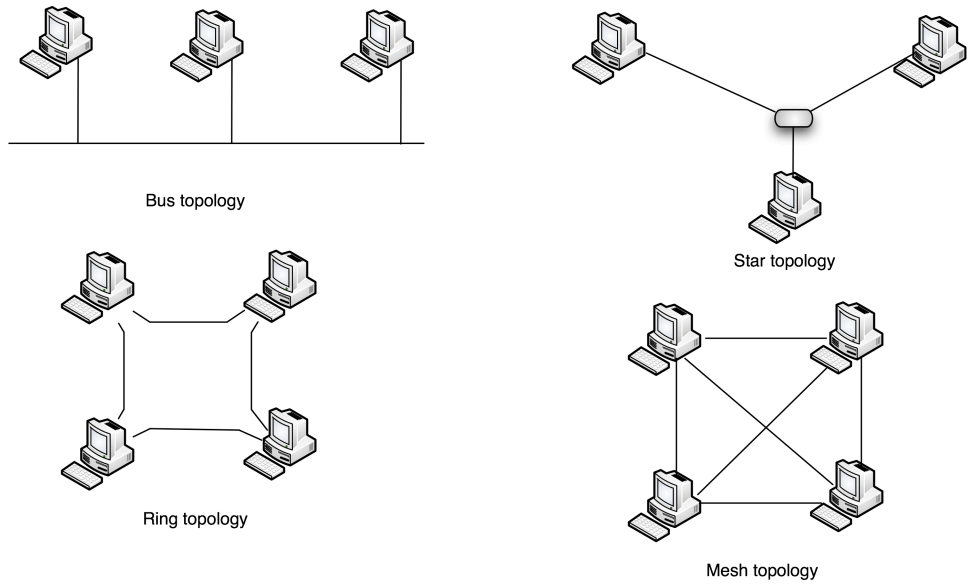

Figure 28: Basic network topology types based on [Bos07, pp. 71-73]

Selected characteristics four basic topology types according to [Bos07, pp. 71-73] are:

- Bus topology (linear bus, single cable to which all nodes are connected using short cables),
- Star topology (contains a main node to which all other nodes are connected with a single connection),
- Ring topology (each node is connected to two neighbours creating a closed ring),
- Mesh topology (each node is connected to one or more other nodes).

This basic topology and is actual implementation is one of the characteristics, which can yield important information for network forensics (see Section 4.2.1.3.2).

The communication in networks in popular use for IT systems is standardised. Nearly all communication systems can be described to a large degree using the Open System Interconnect (OSI) model, which is standardised by the International Organization for Standardization (ISO) as ISO/IEC 7498 and is commonly known as the OSI/ISO reference model (see [Cas11, p. 621] based on [ISO19]). We will use this existing model to structure data organisation with regards to layers of communication.

Although the widely used and thus highly relevant TCP/IP protocol suite is a concurrent development (see [Rus19]), its layered structure can be described using the OSI/ISO reference model [Cas11, p. 621]. The following Figure 29 shows the general structure of the data organisation in networks when applying the ISO/OSI reference model (see also [Cas11, p. 629]).

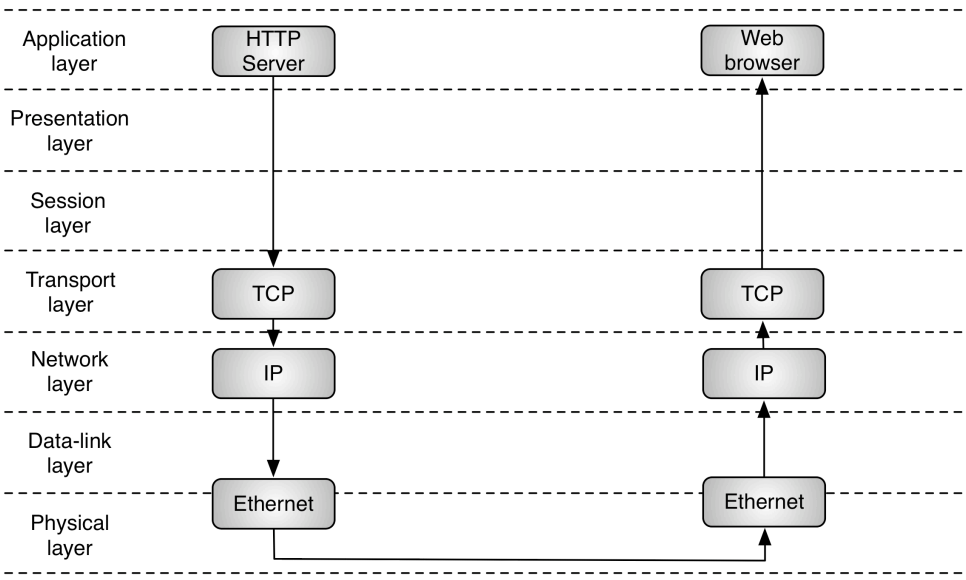

Figure 29: Generalised structure of data organisation in networks using the OSI/ISO reference model on the example of a http server sending data packets to the web browser client system applying the TCP/IP protocol (modified from [Cas11, 629])

As can be seen from that figure, not necessarily all OSI layers have to be present in a given protocol. The layering of TCP/IP is described in [Cas11, p. 629]) as follows:

- Physical layer: the actual media that carries the data (e.g. telephone wires, fibre optics, radio signals etc.),
- Data-link layer: provides basic connectivity between computers close to one another, basically the data formation for the processing by computers and transportation by the respective medium,
- Network layer: responsible for routing data to its destination using addresses (data delivery),
- Transport layer: establishes, maintains, manages, and terminates communications between hosts, divides large messages into smaller, more manageable parts and reconstruction on reception at the other end,
- Session layer: coordinates dialog between hosts, establishing, maintaining, managing, and terminating communications,
- Presentation layer: formats and converts data to meet the conventions of the specific computer being used,
- Application layer: provides the interface between people and networks.

In general, network communication is typically structured into protocols, which are essential for its proper function and as such are a source of forensically essential information (see Section 2.7.2), as the network communication could not be established without adhering to (portions of) the protocol.

In the following we will select some exemplarily chosen characteristics of network management for forensics.

4.2.1.3.2 Useful characteristics of network management for forensics and mapping to forensic data types

Network forensics is unique compared to mass storage and main memory forensics in that it allows remotely gathering, investigating and analysing potential traces from IT systems. Potential relevant data cannot only be found on the systems triggering the investigation but also in several network management devices (e.g. switches, router, firewalls).

During strategic preparation (see Section 4.4.1.1) the operator of an IT system can execute a number of measures that can greatly support a forensic examination. At the very least, a network plan (supporting a digital evidence map from [Cas11, p. 643] as well as a system landscape analysis from Section 4.4.1.1) can be made available. It contains the actual network topology (see Section 4.2.1.3.1) and reflects the most recent system layout, based on the networks topology, i.e. structure consisting of network nodes and connections (see e.g. [Bos07, p. 71]). This allows for a deeper interpretation of communication relations derived from some of the nodes.

Of even more benefit is the placement of physical access points to the network that only serve the forensic acquisition of network traffic data and thus need to be designed least intrusive with regards to the functional network traffic. Otherwise not only the network traffic will be altered, further an attacker could be alerted to the presence of the examiners and take evasive or destructive actions. Since the introduction and widespread use of network switches (both standalone or as part of router devices), the facility to monitor an arbitrary point in the network requires additional effort. Some switches offer the establishment of a monitor port, also known as a Switched Port Analyzer (SPAN) port (see e.g. [LGS+10, p. 50]. This allows monitoring network traffic flowing through the port connections to be mirrored on the monitor port (see Figure 30).

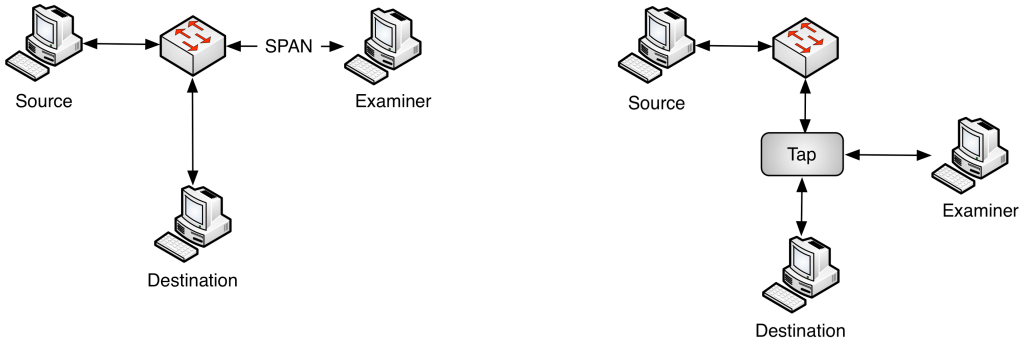

Figure 30: Network traffic acquisition using SPAN ports and network taps (modified from [LGS+10, p. 50]) as a useful characteristic of network management for forensics

However, as stated in [LGS+10, p. 50], this has some drawbacks (e.g. alteration of packet timing, frame drops because of device overload during the SPAN process, cleaning of traffic and dropping of corrupt packets). The introduction of an additional piece of hardware known as a network tap, which is completely passive with regards to the original network can alleviate the problems of SPAN ports. Using SPAN ports and network taps and acquisition devices, ideally with the Linux Transparent Forensic Bridge (LFTB) [KHA+10, pp. 96-99] outlined in Section 4.4.1.1 or similar means for forensic soundness, raw network communication data can be acquired. We define any such raw network communication data content as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the following Table 19 shows.

Forensic data type	Description according to [KHD09]	Context in network forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Raw network communication data gathered from strategically chosen vantage points off the communication media (e.g. copper wire, electromagnetic waves etc.)

Table 19: Data gathered from network media assigned as a source for raw data DT_1 as defined in [KHD09, pp. 2-3]

With further processing of this data during the data investigation (Section 4.4.1.4) and data analysis (Section 4.4.1.5) steps, other data potentially types of the network data stream encapsulated therein can be extracted. Such data can be hardware details from the communicating networking elements such as hardware data (e.g. MAC addresses), which we define as hardware data DT_2, according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. Further, with the use of networked file systems such as CIS, NFS), also meta data about the items (e.g. privi-

leges, MAC times) is communicated. We define those meta data as details about data DT_3 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. The network devices can be configured via the network data stream itself (e.g. logging facilities on network devices controlled via SNMP, router/switch port configuration). We define those data as configuration data DT_4 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. But not only the devices themselves, also the network protocol used, can be configured with data on the network stream itself. We define those data as communication protocol data DT_5 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. Data for inter process communication in networked environments (e.g. pipes) is also transferred using the network data stream, which we define as session data DT_6 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. Whole sessions are managed using data transferred on the network stream (e.g. login procedures, cookies). We define them as session data DT_7 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows. Finally, all data a user accesses from networked locations (e.g. media, documents) is transferred on the network data stream, which we define as user data DT_8 according to the data type definition in Table 3 in Section 4.2.1, as Table 20 shows.

Forensic data type	**Description according to [KHD09]**	**Context in network forensics characteristics**
Hardware data (DT_2)	Data not or only in a limited way influenced by the OS and application	Hardware details about the networking equipment necessary for communication (e.g. MAC addresses) over the network data stream
Details about data (DT_3)	Meta data describing other data	Networked filesystems transmit meta data (e.g. permissions, MAC times) over the network data stream
Configuration data (DT_4)	Modify the behaviour of the system and applications	Networking equipment is configured over the network data stream (e.g. physical port configuration)
Communication protocol data (DT_5)	Modify the communication behaviour of the system	The network protocol and configurable properties can be set over the network data stream
Process data (DT_6)	Data about a running process	Inter-process communication for networked devices is exchanged over the network data stream
Session data (DT_7)	Data collected by a system during a session	Whole sessions can be managed over the network data stream (e.g. using cookies)
User data (DT_8)	Contents created, edited or consumed by the user	User accessible/editable content (e.g. media, documents) are transferred over the network data stream

Table 20: Forensic data types encapsulated in network streams using protocols that can be derived from raw data DT_1 as defined in [KHD09, pp. 2-3]

This is down to the fact, that the acquisition does not apply any filtering per se. The collection is most like to occur during live forensics (see Section 3.3.1.2), i.e. with the suspected incident still ongoing.

The devices can employ network-based system management facilities, e.g. by using the simple network managing protocol (SNMP, see [CFS20]) based on UDP packet communication within

the respective network [MaS01a, SNMP and UDP]. SNMP is a client(agent)/server(manager)-based system [MaS01a, SNMP and UDP]. SNMP involves a set operations and the information gathered/manipulated therein aimed at network administrators to enable them to gather information (monitoring) and change aspects of managed network devices [MaS01, What Is SNMP?]. This information can be of high forensic significance and can contain monitored data such as (exemplary and very small selection, see [MaS01a, Table 2-5]):

- system description (hardware data DT_2)
- interfaces description (configuration data DT_4)
- tcp description (communication protocol data DT_5)

However, access to this data requires the cooperation of the administrator and the SNMP service hast to be installed and maintained as part of strategic preparation (see Section 4.4.1.1). Great care has to be taken, that only information gathering takes place and even the action of data gathering can impact the very network communication about to be recorded (see Section 3.3.1.3 for a discussion on the impact of methods for the forensic process)

If access to a *border router* is available, the *routing of the network traffic* can also be examined. According to [LBF+09, p. 173] such a router stands between an Autonomous System (AS, i.e. the local network of an entity) and the Internet Service Provider (ISP, i.e. the entity giving access to a network of networks). Such a border router handles the protocols to the *internal network* (e.g. TCP/IP etc., see Section 4.2.1.3.1) and the protocols to the *external network* (see Figure 31).

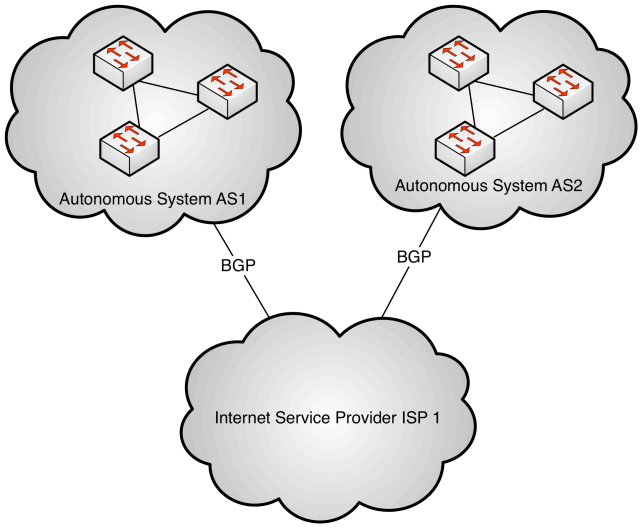

Figure 31: Border gateway to connect exterior and interior networks (as Autonomous Systems AS) using the border gateway protocol (BGP) providing routing information (modified from [LBF+09, p. 174]).

Here, other protocol families such as distance vector routing protocols (communication across whole networks, e.g. Routing Information Protocol RIP, Enhanced Interior Gateway Routing Protocol EIGRP, Border Gateway Protocol BGP) or link state routing protocols (communication to the next hop, e.g. Open Shortest Path First OSPF) are employed (see LBF+09, pp. 172-178). The provide optimal routes from the sending entity to the receiver entity based factors such as hop count, reliability, bandwidth, delay and costs. Additionally, if access to border routers and other information from ISPs exist, also routing information such as hop count towards a subnet and the best route towards the destination network as well as topology updates can be obtained. Thus, from access to border routers and other WAN management facilities, primarily data can be obtained, which we define as communication protocol data DT_5 according to the data type definition in Table 3 in Section 4.2.1, as Table 21 shows.

Forensic data type	Description according to [KHD09]	Context in network forensics characteristics
Communication protocol data (DT_5)	Modify the communication behaviour of the system	The WAN protocol and configurable properties can be set over the network data stream

Table 21: WAN protocol data assigned as communication protocol data DT_5 as defined in [KHD09, pp. 2-3]

The information content is greatly influenced by the choice of the communication protocol that encapsulates data according to semantics. For instance, the TCP/IP protocol (see Section 4.2.1.3.1) provides a richer and pre-structured data that provides an easy mapping to the forensic data types when compared to the CAN bus protocol provided as additional information in the appendix Section 0.

More generally due to the network structure (local as well as wide area), potential case-specific information can be found throughout the whole communication route from sender to receiver collected by the *network devices* used to route the information. Volatile and non-volatile data can be gained from network infrastructure elements (see e.g. [LBF+09, p. 262]) such as (managed) switches (i.e. those having a HW/SW or SW-only interface to the investigator, see e.g. [LBF+09, p. 119]) and routers [LBF+09, p. 119]) as well as hardware-firewalls [LBF+09, p. 126]), intrusion detection/prevention systems and similar network security devices. In some network a whole network security suite is maintained as a Security Event Management System SEM [LBF+09, p. 34], which compiles and correlates logs from various sources in the network. SEM can perform analyses automatically. However, as stated in [LBF+09, p. 34], the original logs need to be kept, since the aggregation is lossy and the automated analysis might miss certain events. Generally, network (hardware) devices could contain a number of potential case-specific forensic data and, if accessible, can be (up to a point) subjected to mass storage stream (typically flash-based storage) forensics (see Section 4.2.1.1.2) and even main memory (typically conventional RAM) forensics (see Section 4.2.1.2.2), especially if conventional desktop IT devices (i.e. "beige" hardware) are used to provide the functionality.

In the next section we will look into the data types for digitised forensics, which are arranged quite differently but following the same characteristics as laid out in the beginning of Section 4.2 but embedded into the context of digitised forensics.

4.2.2 Data types for digitised forensics

We use our existing findings from [KDV12, pp. 294-299] presented to the English speaking audience in [KDV15, p. 89] and extend them to the data stream idea, where all of the forensic data types form a common forensic data source and check their applicability in this new context. Also we check whether the definitions need updating. We use M2 of our methodology (Section 1.2) and achieve the contribution C2.2 (Section 1.3).

The data types for digitised forensics also form a data source similar to the one devised for digital forensics. However, the input source is different as typically contact-less sensors S_1 to S_n (to bring up the advantages of digitised forensics, see Section 3.4) provide the data stream $DS_{S1...Sn}$ by acquiring selected aspects of the physical environment and represent them as digital data. All data streams form the data source for digitised forensics.

Note: This thesis only addresses digital data, for a detailed discussion of the boundary between the physical world and the digital domain the reader is directed to [Hil20].

The following Figure 32 depicts the data source for digitised forensics.

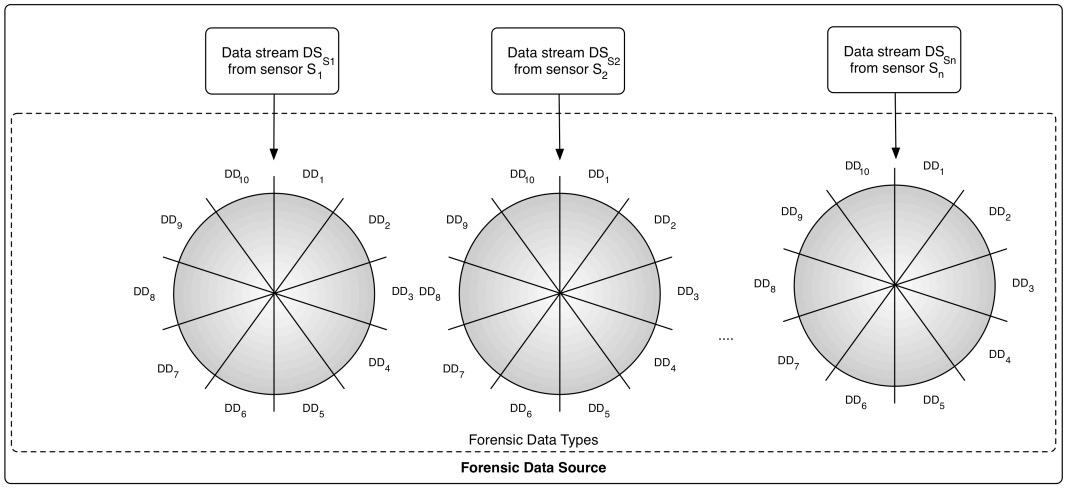

Figure 32: Forensic Data Source separated into data streams for digitised forensics SF

As the Figure 32 indicates, the forensic data source includes data acquired from all data streams based on the idea that acquiring, investigating and analysing data from different sensors and their the correlation and/or fusing of data streams from different sensors is likely to yield better results of better evidentiary value (see also Section 2.7.2). However, based on availability sometimes only one sensor is used.

As stated in Section 4.2 we apply the following characteristics embodied in the ISO/OSI model of the ISO/IEC 7498 ([Cas11, p. 621] based on [ISO19]):

- layering of data by giving them semantics with regards to a specific purpose,
- layering of data in a way that the same semantics can be present in different syntactical encapsulations.

The layering (as proposed in [KDV12, pp. 294-299]) is modelled alongside the biometric pipeline (see Section 2.8 and the signal processing and pattern recognition pipeline from Section 3.4 derived thereof.

The general idea as introduced by [KDV12, pp. 294-299] and presented to the English speaking audience in [KDV15, p. 89] is to model the forensic data types for digitised forensics alongside the biometric pipeline for biometric user authentication (see Section 2.8). The idea is to encapsulate the different steps necessary to process this pipeline since the data is likely to be processed in a similar manner if it belongs to the same step. Analogous to the ISO/OSI model [ISO19], the layering is not mutual. Same as with the layering for digital forensics DF, raw data forms the basis of the data streams generated by the different sensors, out of which the other forensic data types DD for digitised forensics SF can be extracted, which encapsulate similar properties and means of processing in the forensic process. The following Figure 33 depicts the encapsulation according to the pipeline. The parameter data DD_4 deserve a special mention as they are introduced into the process by the examiner and are not part of the trace object itself. However, many surface measurement systems used in digitised forensics embed the parameter data into the raw sensor data. Thus, for all intents and purposes of this thesis, parameter data are part of the forensic process in digitised forensics and can be used, at least in principle, throughout the whole signal processing and pattern recognition pipeline.

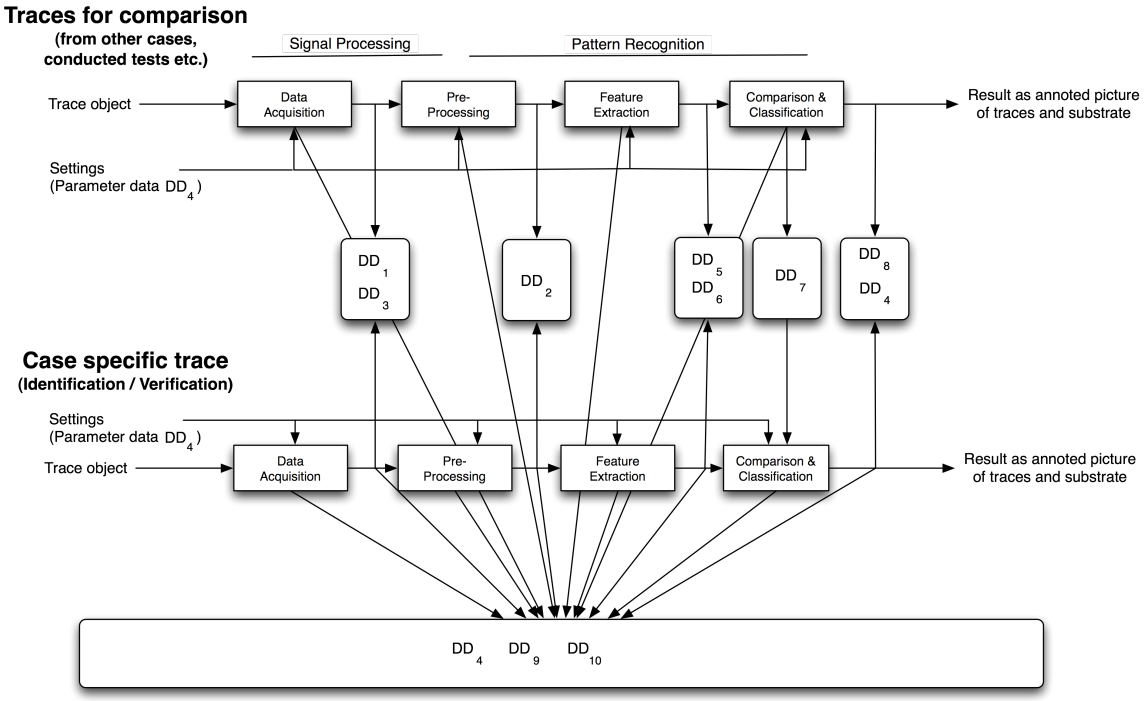

Figure 33: Forensic data types for digitised forensics as they are used alongside the signal processing and pattern recognition pipeline from Section3.4

In the following we explain the forensic data types DD depicted in Figure 33 after summarising them in the following Table 22 as proposed to the English-speaking audience in [KDV15, p. 89]:

Forensic data type	**Description according to [KVD15]**
Raw sensor data (DD_1)	Digital input data from the digitalisation process (e.g. scans of test samples)
Processed signal data (DD_2)	Contain the results of transformations to raw sensor data (e.g. visibility enhanced fingerprint pattern)
Contextual data (DD_3)	Contain environmental data e.g. spatial information, spatial relation between traces, temperature, humidity
Parameter data (DD_4)	Contain settings and other parameter used for acquisition, investigation and analysis
Trace characteristic feature data (DD_5)	Describe trace specific investigation results, e.g. level1/2/3 fingerprint features
Substrate characteristic feature data (DD_6)	Describe trace carrier specific investigation results, e.g. surface type, individual surface characteristics
Model data (DD_7)	Describe trained model data, e.g. surface specific scanner settings, reference data
Classification result data (DD_8)	Describes classification results gained by applying machine learning and comparable approaches
Chain of custody data (DD_9)	Describe data used to ensure integrity and authenticity and process accompanying documentation, e.g. cryptographic hash sums, certificates, device identification, time stamps
Report data (DD_{10})	Describe data for the process accompanying documentation and for the final report

Table 22: Forensic data types DD for digitised forensics SF taken from [KDV15, p. 89]

The pipeline starts with the acquisition of a selection of physical aspects of a trace and trace carrier. Here we denote a slight deviation from the biometrics pipeline as described in Section 2.8 as we do not acquire behavioural/physiological traits but (typically latent) traces left at a crime scene by an individual (see also Section 3.4)

The following description of forensic data types DD for digitised forensics SF is based on [KDV12, pp. 294-299] and has been translated into the English language and adapted in terms of wording and sequencing according to Table 22 and Figure 32. We start with the data stream that is supplied by the sensor in its untreated, raw form, i.e. the first emergence of digital data from the digitised aspects of physical objects together with a substrate, which we define as raw sensor data DD_1 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 23 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Raw sensor data (DD_1)	Digital input data from the digitalisation process (e.g. scans of test samples)	Result of the digitalisation of aspects of physical traces and the substrates they are contained on/in

Table 23: Forensic data type as result of the digitalisation assigned as raw data DD_1 as defined in [KDV15, p. 89]

The actual result of running through the whole pipeline is the image of the trace removed from its substrate and, depending on the respective use case/subtask of pattern recognition annotated with features relevant for the forensic examiner. We define this forensic data type data as processed signal data DD_2 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 24 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Processed signal data (DD_2)	Contain the results of transformations to raw sensor data (e.g. visibility enhanced fingerprint pattern)	Enhanced/annotated trace depiction for presentation towards the forensic experts

Table 24: Forensic data type as the result of the pattern recognition pipeline as processed signal data DD_2 as defined in [KDV15, p. 89]

Depending on the sensor used, further data containing contextual information of the scan process can be embedded, which enrich the data with information about the acquisition environment such as spatial information, scales used during the digitalisation process etc. In many senses, this is the equivalent of meta data or details about data in digital forensics (see Section 4.2.1). We define this forensic data type as contextual data DD_3 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 25 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Contextual data (DD_3)	Contain environmental data e.g. spatial information, spatial relation between traces, temperature, humidity	Meta data of the digitisation process recorded by the acquisition device

Table 25: Forensic data type as meta data of the digitalisation process as contextual data DD_3 as defined in [KDV15, p. 89]

The acquisition devices and the software used in the pattern recognition pipeline are often highly configurable. By setting those options to specific values, configuration data are created, again quite similar to configuration data used in digital forensics (see Section 4.2.1). Thanks to the discussion in the Ph.D. defence with the thesis advisor Jana Dittmann, this data type forms a connection to the configuration data DT4 for digital forensics. We argue that we can define a relation where set of parameter data for digitised forensics DD_4 is smaller in their size than the configuration data DT_4 for digital forensics on the grounds that are much more configurable options on IT systems used as measurement support systems for digitised forensics than just scan parameters. We define this data type as parameter data DD_4 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 26 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Parameter data (DD_4)	Contain settings and other parameter used for acquisition, investigation and analysis	Configuration data of the acquisition devices and all other configurable options of the pattern recognition pipeline

Table 26: Forensic data type as configuration data throughout the pipeline as parameter data DD_4 as defined in [KDV15, p. 89]

Trace type specific feature data act as input for the pattern recognition pipeline. We define this forensic data type as trace characteristic feature data DD_5 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 27 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Trace characteristic feature data (DD_5)	Describe trace specific investigation results, e.g. level1/2/3 fingerprint features	Input for pattern classification performed on the separated trace data used to annotate the final result in aid of the forensic expert

Table 27: Forensic data type as input for pattern classification as trace characteristic feature data DD_5 as defined in [KDV15, p. 89]

Substrate type specific feature data act as input for the pattern recognition pipeline (sometimes, the substrate itself is subject to evaluation, see also Section 3.4.1). We define this forensic data type as substrate characteristic feature data DD_6 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 28 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Substrate characteristic feature data (DD_6)	Describe trace carrier specific investigation results, e.g. surface type, individual surface characteristics	Input for pattern classification performed on the separated substrate data used to annotate the final result in aid of the forensic expert

Table 28: Forensic data type as input for pattern classification as substrate characteristic feature data DD_6 as defined in [KDV15, p. 89]

The models used in the pattern recognition pipeline need to trained ahead of a specific case (strategic preparation, see Section 4.4.2.1), the data used from those training purposes we define as model data DD_7 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 29 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Model data (DD_7)	Describe trained model data, e.g. surface specific scanner settings, reference data	Model training needs to occur ahead of a specific case to enable pattern recognition functionality

Table 29: Forensic data type as supplied model data from the strategic preparation as model data DD_7 as defined in [KDV15, p. 89]

The results of the pattern classification pipeline applied using both trace characteristic feature data DD_5 and substrate characteristic feature data DD_6 and processing raw signal data DD_1 we define as classification result data DD_8 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 30 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Classification result data (DD_8)	Describes classification results gained by applying machine learning and comparable approaches	Results of pattern classification performed on the separated trace and substrate data used to annotate the final result in aid of the forensic expert

Table 30: Forensic data type containing pattern classification results as classification result data DD_8 as defined in [KDV15, p. 89]

Throughout the whole digitised forensics process, the maintenance of the IT security aspects (primarily integrity, authenticity and non-repudiation) needs to be documented and verified. This is ever more important because the all digital processing of latent data with no visible clue towards trace shape and position needs other means to ensure comprehensibility (see Section 3.4). All data created using digital means we define as chain of custody data DD_9 according to the data type definition in Table 22 in Section 4.2.1 as the following Table 31 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Chain of custody data (DD_9)	Describe data used to ensure integrity and authenticity and process accompanying documentation, e.g. cryptographic hash sums, certificates, device identification, time stamps	Maintenance of IT security aspects throughout the whole digitised forensic process for documentation and verification purposes

Table 31: Forensic data type for maintenance of IT security aspects as chain of custody data DD_9 as defined in [KDV15, p. 89]

The results of the (partially) automated digitised forensic examination and the results from the forensic experts together with their conclusions based on the annotated trace representation need to be forged into a conclusive final report. All digital data used for compiling this report we define as the forensic data type of report data DD_{10} according to the data type definition in Table 22 in Section 4.2.1 as the following Table 32 shows:

Forensic data type	Description according to [KVD15]	Context to the digital forensic process
Report data (DD_{10})	Describe data for the process accompanying documentation and for the final report	Contains all digital data used by the forensic expert to conclude the final decision based on the annotated depiction of trace and substrate

Table 32: Forensic data type for the final report as report data DD_{10} as defined in [KDV15, p. 89]

It can be argued that from the perspective of digital forensics DF regarding the IT systems involved (see Section 4.2.1), all forensic data types of digitised forensics represent DT_8 (user data). This becomes particularly relevant and offers support when looking into an incident where an IT system used for digitised forensics has been attacked and requires an examination according to digital forensics.

4.2.3 Reflection and parting thoughts

We individually applied M2 as part of our methodology (see Section 1.2) to the specific application areas of digital forensics DF and digitised forensics SF, respectively.

> M2:
>
> We apply the selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 261] to data stored, processed, communicated in IT systems to distinguish forensic data types for digital and digitised forensics. To achieve this, we construct layers of data by giving them semantics that support the forensic process. In further conjunction with the ISO/OSI model, we use a layering that is not mutual exclusive.

We looked into the data streams in IT systems used in digital forensics DF and digitised forensics SF and established a layering based on semantics that can support the forensic process as contribution C2 (including C2.1 and C2.2).

> C2:
> Establishment of a data-centric (data-driven) view in digital and digitised forensics by applying a structured layering of data based on selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11] in various IT systems (from embedded systems/IoT to data centres) to formulate forensic data types:
> C2.1:
> Adding semantics to data in the forensic context and establishment of 8 data types for digital forensics for selected use cases
> C2.2:
> Adding semantics to data in the forensic context and establishment of 10 data types for digitised forensics for selected use cases

By encapsulating the forensic data types into a layered, non-mutual exclusive structure we found a means to abstractly describe their essence and importance for forensics without references to specific products and manufacturers.

In the following sections we look into the data processing facilities that operate on the forensic data types DT for digital forensics DF and DD for digitised forensics SF as investigation target D_{IT} and the investigation result D_{IR} and structure the characteristics of those facilities.

4.3 Sets of methods for the forensic process

In Section 4.2 we introduced the connection between the estimation of loss, error and uncertainty in forensic examinations and the forensic data those examinations operate upon as the investigation target D_{IT} and the investigation result D_{IR} as elements of the forensic data types DT and DD for digital and digitised forensics, respectively. To achieve this, in Section 4.1 we defined both the data forming the investigation target D_{IT} and the investigation result D_{IR} as a subset of the set of data containing information D_I. We now look into the facilities that transform the data forming the investigation target D_{IT} into the investigation result D_{IR}.

The choices of those facilities have a profound impact on the results (see e.g. Section 5). We thus recall the research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

We now look into the preservation of tool sovereignty by formalising the processing facilities and discovering shared properties, which allow a grouping based on similarities in those properties. In informal starting point and attempt to categorise sets of methods is our work in [KHD09, p. 2], which we extend in the following.

To describe the properties of those processing facilities, we use the transfer functions mentioned in Section 4.1 as equation 4.16 based on [Car03, p. 4]:

$$f_{DF,SF}\,(d_{IT}) : D_{IT} \longrightarrow D_{IR} \qquad (4.16)$$

that are designed to describe tools and methods according to the transformation of data from the investigation target D_{IT} into the investigation result D_{IR}.

We formally define those processing facilities formed by forensic tools and methods as *methods of the forensic process* as follows:

> Methods of the forensic process:
>
> A method of the forensic process is a set of forensic transfer functions that are executed by entities with common properties. An entity is a piece of code with useful characteristics for the forensic process and is involved in either the detection, acquisition, investigation, analysis and/or documentation activities. Methods of the forensic process use transfer functions (modified from [Car03, p.4]) operating on the data of the investigation target D_{IT} and produce investigation result data D_{IR}. Sets of methods of the forensic process are mutually exclusive and thus form a residual class structure.

Looking into shared properties of sets of methods for the forensic process follows the *methodology item M3* from Section 1.2:

> M3:
>
> We use residual class based hierarchical approaches (as opposed to a layered non- mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, p.4]. We further use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics.

Sets of methods for the forensic process are based on transfer functions [Car03, p. 4] but they need some alterations before usage in the DCEA approach introduced in this thesis. The following Figure 34 contains our additions and alterations.

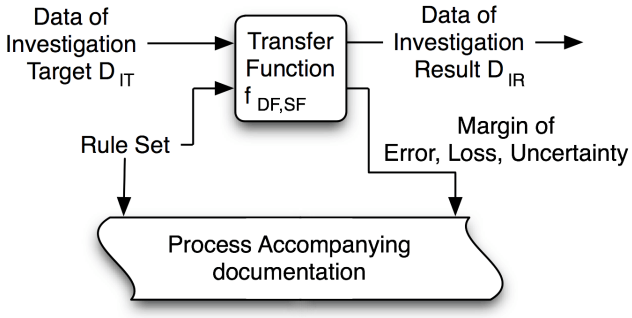

Figure 34: Transfer function, derived and enhanced from [Car03, p. 4]

As stated in Section 2.4 and re-iterated in Section 4.4, the documentation of all proceedings together with the *data of the investigation target* D_{IT}, *data of the investigation result* D_{IR} and its configuration is vital. From strategic preparation SP with included forensic tool testing and benchmarking (see Sections 4.4.1.1 and 4.4.2.1) and from past experiences with forensic examinations, the *rule set* is derived. On the output side we not only look at *errors* but also *loss and uncertainty*. So we alter the depiction for input, output, function and margins and add the process accompanying documentation accordingly.

To incorporate sets of methods, we connect a multitude of methods according to the layering of the forensic data type used and the data stream they operate on. The following Figure 35 shows this connection generically for both digital forensics DF and digitised forensics SF.

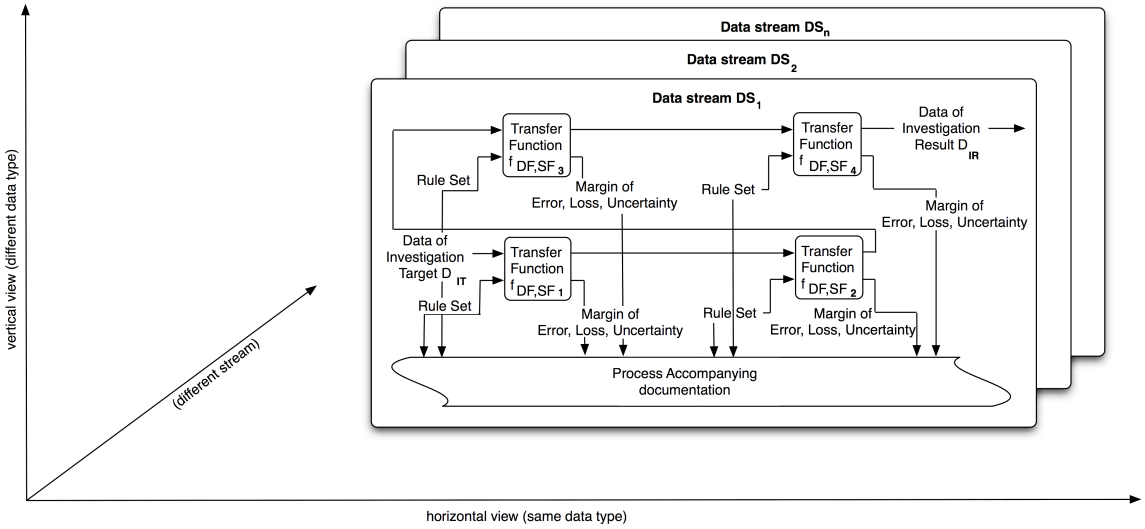

Figure 35: Transfer functions enhanced from [Car03, p. 4] in the context of layers and data streams for digital forensics DF and digitised forensics SF

The data of the investigation result D_{IR} as a result of a transfer function operating on the data of the investigation target D_{IT} can be of the same forensic data type DT, DD or of a different data type. Additionally, those operations can take place in the different data streams DS. Obviously, in digital forensics, only three data streams have been identified where in digitised forensics the number of data streams depends on the number of sensors used.

Methods for the forensic process changing forensic data types are somewhat similar to the layer translation (abstraction layer) from [Car03, p. 3]. They can incorporate an intended reduction of data (similar to the lossy translation in [Car03, p. 5]).

More formally, we describe a forensic examination in digital forensics DF and in digitised forensic SF as a concatenation of sets of methods for the forensic process in all data streams DS as follows (equation 4.17):

$$F_{DF,SF}^{DS_1} \times F_{DF,SF}^{DS_2} \ldots \times F_{DF,SF}^{DS_n} \quad (4.17)$$

However, the trivial solution is also valid, i.e. a whole examination can work on a single layer exclusively.
By defining sets of methods we address the demand from [CaS03, p. 1] for generality and independence from current products and procedures. They represent the computer scientist's view according to [PBM08, p. 114]. Since the underlying principles for the categorisation differs for digital forensics DF and digitised forensics SF, we address them in different subsections. We start with sets of methods for the digital forensics process in the following Sections 4.3.1 and continue with sets of methods for the digitised forensics process in Section 4.3.2.

4.3.1 Sets of methods for digital forensics

As stated in Section 4.3, the sets of methods for the forensic process are designed to be mutually exclusive and forming a residual class structure. We introduce a grouping based on the practicalities of the forensic process.

The examiner looks for electronic traces by looking at an abstraction of the services used to fulfil the tasks from the user perspective, which can also have a forensic functionality/side use (e.g. by logging data or arranging data in a particular way that can be exploited forensically).

We order the sets of methods for the forensic process according to an *estimation of availability* to the forensic examiner from most common to rarest based on the definitions in [KDV15, p. 88] based on the work from [KHD09, p. 2] as summarised in Table 33.

Sets of methods for the forensic process in digital forensics	**Description according to [KDV15]**
Operating system (OS)	Methods that provide forensically relevant data besides serving their main purpose of distributing computing resources
File system (FS)	Methods that provide forensically relevant data besides serving their main purpose of maintaining the file system
IT application (ITA)	Methods provided by IT applications that provide forensically relevant data besides serving their main purpose
Explicit means of intrusion detection (EMID)	Methods that are executed autonomous on a routine basis and without a suspicion of an incident
Scaling of methods for evidence gathering (SMG)	Methods that are unsuited for routine usage in a production environment e.g. because of false positives, high CPU demands)
Data processing and evaluation (DPE)	Dedicated methods of the forensic process that display, process or document information

Table 33: Grouping of sets of methods for the forensic process in digital forensics based on an estimation of availability to the forensic examiner (from most common to rarest), based on the content definitions from [KDV15, p. 88]

Operating systems as resource managers in some shape or form are likely to be present in any device implementing a random access stored program (RASP) architecture (see [Har71, pp. 234-240]). The forensic value of operating systems is enormous since it has some access to all facilities of the IT system (often with the highest privileges). The operating system often manages the memory subsystem and thus contains more volatile data (see Section 3.3.1.1). However, a notable

exception is the virtual memory system (see Section 4.2.1.2.1). The non-functional (i.e. not providing a service to the normal users of the IT system) aspects used by digital forensics and forming the set of methods for forensics of the operating system OS include log keeping and the detection of privilege escalation. The following Table 34 (based on Table 33) summarises the main characteristics of the sets of methods for the operating system OS and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
Operating system (OS)	Methods that provide forensically relevant data besides serving their main purpose of distributing computing resources	Resource managing OS with broad access to data in the IT system, mostly higher volatile data, extensive log keeping

Table 34: Characteristics of the sets of methods of the operation system OS in digital forensics DF based on [KDV15, p. 88]

File systems primarily operate on mass storage devices (see Section 4.2.1.1) and thus the mass storage data stream DS_T. They manage data with a lower volatility (see Section 3.3.1.1). Nearly all IT systems use a filing mechanism of some sort (including data base system based variants) in order to store and retrieve data not in permanent use. File system forensics is one of the oldest and very well researched areas of digital forensics; whole books focus on the forensics of file system (e.g. [FaV05]). Aside from the service provided by the intended user, its non-functional properties used in digital forensics as the set of methods for forensics from the file system FS include the maintenance of meta data (permissions, MAC times etc.) and log keeping (e.g. file system journals). The following Table 35 (based on Table 33) summarises main characteristics of the sets of methods for the file system FS and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
File system (FS)	Methods that provide forensically relevant data besides serving their main purpose of maintaining the file system	Maintenance of meta data and log keeping, typically operating on low-volatility storage

Table 35: Characteristics of the sets of methods of the file system FS in digital forensics DF based on [KDV15, p. 88]

IT applications are the sole reason for using IT systems in the first place as they provide functionality to the user of the IT system. They are thus very likely to be available on an IT system. They can operate on all data streams DS of an IT system: mass storage DS_T (see Section 4.2.1.1), main memory DS_M (see Section 4.2.1.2) and network DS_N (see Section 4.2.1.3) and cover different levels of volatility from low (e.g. files in the file systems) to very high (data maintained in main memory or communicated over the network). The non-functional properties used in digital forensics as a set of methods for forensics from the IT application ITA include session management, log keeping, storage and communication of user contents enriched with meta data etc. Application forensics (e.g. in database management systems or word processors) is already quite advanced (see e.g. SAP ERP Forensics [Shi16, pp. 16-22] for a discussion of builtin logging mechanisms) etc. Table 36 (based on Table 33) summarises the main characteristics of the sets of methods for the IT application ITA and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
IT application (ITA)	Methods provided by IT applications that provide forensically relevant data besides serving their main purpose	Can operate on all data streams, different levels of volatility, session management, log keeping, storage and communication of user content

Table 36: Characteristics of the sets of methods of the IT application ITA in digital forensics DF based on [KDV15, p. 88]

There are uses of IT systems that are third party (i.e. not operating system or file system) and still rather support the maintenance of the system and are often hidden away from the user of an IT system in normal circumstances. Since they pose so little visible use to the end user of an IT system, they are less likely to be found on IT systems. In [KHD09, p. 2] they are introduced as explicit means of intrusion detection and operate on all data streams DS of an IT system: mass storage DS_T (see Section 4.2.1.1), main memory DS_M (see Section 4.2.1.2) and network DS_N (see Section 4.2.1.3) and cover different levels of volatility from low (e.g. files in the file systems) to very high (data maintained in main memory or communicated over the network). In [HKD09, p. 234] an approach for usage of IDS we also apply them to the non-standard application field out automotive systems and suggest a support for IT forensics based on the anomalies detected. Those programs and services (e.g. virus protection software, intrusion detection systems etc.) are resources savvy and routinely executed in the background. Apart from its intended functional use, its non-functional properties used in digital forensics as a set of methods for forensics from the explicit means of intrusion detection EMID include log keeping and record generation/maintenance. The following Table 37 (based on Table 33) summarises the main characteristics of the set of methods for the explicit means of intrusion detection EMID and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
Explicit means of intrusion detection (EMID)	Methods that are executed autonomous on a routine basis and without a suspicion of an incident	Operate on all data streams, different levels of volatility, routinely execution in the background, log keeping and record generation/maintenance

Table 37: Characteristics of the method for the forensic process of explicit means of intrusion detection EMID in digital forensics DF based on [KDV15, p. 88]

Similar to the programs and services that are third party (i.e. not operating system or file system) and still rather support the maintenance of the system and are often hidden away from the user of an IT system in normal circumstances, programs and services exist that can drain resources and are thus unfit to be executed routinely. Due to their resource demands and little visibility and use to then end user in normal circumstances, they are less likely to be found than EMID methods. However, if a suspicion is raised that the IT system is operating in an abnormal mode (regardless of intentionally maliciously attacked or due to hardware/software failures), they can be invoked and provide the users with details about the abnormal behaviour (e.g. debuggers, on demand virus scanners etc.). In [KHD09, p. 2] they are introduced as scaling of methods for evidence gathering and operate on all data streams DS of an IT system: mass storage DS_T (see Section 4.2.1.1), main memory DS_M (see Section 4.2.1.2) and network DS_N (see Section 4.2.1.3) and cover different levels of volatility from low (e.g. files in the file systems) to very high (data maintained in main memory or communicated over the network). Apart from its intended functional use, its non-

functional properties used in digital forensics as a set of methods for forensics from scaling of methods for evidence gathering SMG include memory dumps, logs for anomaly detection, logs of detected malicious software etc. The following Table 38 (based on Table 33) summarises the main characteristics of the sets of methods for the operating system OS and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
Scaling of methods for evidence gathering (SMG)	Methods that are unsuited for routine usage in a production environment e.g. because of false positives, high CPU demands)	High resource demands, anomaly detection, user invoked activation, memory dump generation, log keeping

Table 38: Characteristics of the sets of methods of scaling of methods for evidence gathering SMG in digital forensics DF based on [KDV15, p. 88]

Finally there are programs and services that serve no other purpose other than direct support of forensic examinations by gathering, investigating, analysing or documenting the data. They are very unlikely to be present on IT systems. By their very nature they operate on all data streams DS of an IT system: mass storage DS_T (see Section 4.2.1.1), main memory DS_M (see Section 4.2.1.2) and network DS_N (see Section 4.2.1.3) and cover different levels of volatility from low (e.g. files in the file systems) to very high (data maintained in main memory or communicated over the network). In [KHD09, p. 2] they are introduced as data processing and evaluation. Its functional properties used in digital forensics as a set of methods for forensics from data processing and evaluation contain all mechanisms known to digital forensics (e.g. image creation, file system investigation, network analysis etc.). The following Table 39 summarises based on Table 33 the main characteristics of the set of methods for the data processing and evaluation DPE and adds context for the forensic process in digital forensics DF.

Sets of methods for the forensic process in digital forensics	Description according to [KDV15]	Context to the forensic process in digital forensics DF
Data processing and evaluation (DPE)	Dedicated methods of the forensic process that display, process or document information	Purpose built for forensics, access to all data streams, all mechanisms known to forensics

Table 39: Characteristics of the sets of methods of data processing and evaluation in digital forensics DF based on [KDV15, p. 88]

In the following we apply the same principle of mutually exclusive grouping of sets of methods for the forensic process for the specifics of digitised forensics.

4.3.2 Sets of methods for digitised forensics

As stated in Section 4.3, the sets of methods for the forensic process are designed to be mutually exclusive and forming a residual class structure. We introduce a grouping based on the practicalities of the forensic process. The categorisation for digitised forensics is based on the usage of IT systems implementing signal processing and pattern recognition (see Section 3.4) The sets of methods as summarised in Table 40 are based on [KDV15, p. 89].

Sets of methods for the forensic process in digitised forensics	Description according to [KDV15]
Methods for digitalisation of physical trace objects (DPO)	Contains methods that transfer physical trace information into digital data, setting the beginning of the digital object chain of custody
Methods for digitalisation of contextual information (DCI)	Contains methods that transfer physical contextual information into digital data, e.g. environmental data
Methods for image enhancement (IE)	Contains methods used in pre-processing to enhance the contrast between trace and substrate surface
Methods for model generation (MG)	Contains methods used to build models for use in machine learning and classification
Methods for feature extraction (FE)	Contains methods that usually automatically extract feature data from traces and substrates
Methods for classification (CL)	Contains methods that use extracted features as input and perform a, typically autonomous classification based on trained model data
Methods for parameter extraction (PE)	Contains methods that derive parameters, e.g. from the classification, to be used for optimised scanning and filter operations
Methods for data fusion (DFU)	Contains methods to combine gathered and pre-processed data, possibly originating from different sensors and or different pre-processing approaches
Methods for presentation and annotation of evidence (PA)	Contains methods that support the manual work of forensic experts by highlighting important trace features
Methods for chain of custody maintenance (CC)	Contains methods that support ensuring the security aspects of integrity and authenticity as well as the process accompanying documentation, e.g. used sensor or filter settings

Table 40: Grouping of sets of methods for the forensic process in digital forensics based on an encapsulation of the pipelined process from Section 3.4 using signal processing and pattern recognition using the content definitions from [KDV15, p. 89] with alteration of abbreviation for methods for data fusion to DFU

The start of the pipeline is the employment of methods for digitalisation of aspects of the physical world, mainly the suspected traces and the substrate's surface carrying them. Those methods inject digital data into the processing pipeline and are the bases of all following data transformations. They also initiate the necessity for a digital chain of custody (see Section 2.4). The digitalisation methods for substrates containing traces for use in digitised forensics SF we define as a set of methods for the digitalisation of physical objects DPO. The following Table 41 (based on Table 40) summarises the main characteristics of the set of methods for digitalisation of physical trace objects DPO and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	Description according to [KDV15]	Context to the forensic process in digitised forensics SF
Methods for digitalisation of physical trace objects (DPO)	Contains methods that transfer physical trace information into digital data, setting the beginning of the digital object chain of custody	Injection of digital data based on sensed physical properties, initiation of digital chain of custody

Table 41: Characteristics of the sets of methods for the digitalisation of physical objects DPO in digitised forensics SF based on [KDV15, p. 89]

Similarly, we use methods that digitise aspects of the environmental conditions and other contextual information (e.g. spatial relation information) during the substrate digitalisation process as meta data. The digitalisation methods for meta information for use in digitised forensics SF we define as a set of methods for digitalisation of contextual information (DCI). Table 42 (based on Table 40) summarises the main characteristics of the set of methods for digitalisation of contextual information DCI and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for digitalisation of contextual information (DCI)	Contains methods that transfer physical contextual information into digital data, e.g. environmental data	Digitalisation of meta data regarding the scan process during the substrate digitalisation process

Table 42: Characteristics of the sets of methods for the digitalisation of contextual information DCI in digitised forensics SF based on [KDV15, p. 89]

Mostly signal processing methods for the processing of the digitised data from the substrate containing suspected traces are employed to enhance the contrast between the substrate's surface and the traces themselves. Those processing methods operating on digitised substrate information for use in digitised forensics SF we define as a set of methods for image enhancement IE. The following Table 43 (based on Table 40) summarises the main characteristics of the set of methods for image enhancement IE and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for image enhancement (IE)	Contains methods used in pre-processing to enhance the contrast between trace and substrate surface	Signal processing applied to digitised data, operating on scanned substrate data

Table 43: Characteristics of the sets of methods for image enhancement IE in digitised forensics SF based on [KDV15, p. 89]

Before any case-specific examination employing pattern recognition can be initiated, the machine learning has to be initiated by training models able to detect the surface and differentiate it from trace material. This also applies the detection of trace-specific characteristics (e.g. minutia points in fingerprint ridge lines). Those models need to be trained ahead in the step of strategic preparation (see Section 4.4.2.1) employing methods for use in digitised forensics SF we define as a set of methods for model generation MG. The following Table 44 (based on Table 40) summarises the main characteristics of the set of methods for model generation MG and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for model generation (MG)	Contains methods used to build models for use in machine learning and classification	Operating case-independent in strategic preparation for use in pattern recognition / machine learning

Table 44: Characteristics of the sets of methods for model generation MG in digitised forensics SF based on [KDV15, p. 89]

From the digitised data containing substrate surface and trace information features can be derived that allow for subsequent classification steps. Such features can be for instance the surface roughness of the substrate or orientation fields and statistical information (e.g. variance, median). Such methods for use in digitised forensics SF we define as a set of methods for feature extraction FE. The following Table 45 (based on Table 40) summarises the main characteristics of the set of methods for feature extraction FE and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for feature extraction (FE)	Contains methods that usually automatically extract feature data from traces and substrates	Derivation of features for subsequent classification

Table 45: Characteristics of the sets of methods for feature extraction FE in digitised forensics SF based on [KDV15, p.89]

The features extracted using methods for feature extraction FE are used for (typically automated) classification methods based on the trained models available from methods for model generation MG. Classification methods are used in machine learning and pattern classification and for digitised forensics to separate trace from substrate or to distinguish between trace characteristics. The can also be used for the search (localisation) of traces. Those classification methods for use in digitised forensics SF we define as a set of methods for classification CL. The following Table 46 (based on Table 40) summarises the main characteristics of the set of methods for classification CL and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for classification (CL)	Contains methods that use extracted features as input and perform a, typically autonomous classification based on trained model data	Classification problems include traces vs. substrate, trace characteristics, localisation of traces

Table 46: Characteristics of the sets of methods for classification CL in digitised forensics SF based on [KDV15, p.89]

The acquisition/digitalisation devices are highly parameterisable and successive scans with altered settings can yield a self-enhancing mechanism for better digitalisation and filtering, if a pattern recognition approach creates a feedback loop. Methods to achieve those optimised settings and parameters for use in digitised forensics SF we define as a set of methods for parameter extraction PE. The following Table 47 (based on Table 40) summarises the main characteristics of the set of methods for parameter extraction and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for parameter extraction (PE)	Contains methods that derive parameters, e.g. from the classification, to be used for optimised scanning and filter operations	Self enhancing feedback loop of parameter enhancing based on parameter extraction and pattern recognition

Table 47: Characteristics of the sets of methods for parameter extraction PE in digitised forensics SF based on [KDV15, p. 89]

As stated in Section 4.2.2, using multiple sensors can yield better results. In order to achieve better results, the data from the respective sensors can be combined or fused in different stages of the signal processing/pattern recognition pipeline (see e.g. [KHD+12, pp. 1505-1506]). Such methods for use in digitised forensics SF we define as a set of methods for data fusion DFU. The following Table 48 (based on Table 40) summarises the main characteristics of the set of methods for data fusion DF and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for data fusion (DFU)	Contains methods to combine gathered and pre-processed data, possibly originating from different sensors and or different pre-processing approaches	Fusion of different streams at different stages of the signal processing and pattern recognition pipeline

Table 48: Characteristics of the sets of methods for data fusion DF in digitised forensics SF based on [KDV15, p.89]

Since the evaluation of the traces finally rests on a human forensic examiner (see Section 3.4), this expert needs to be presented with a (typically) pictorial representation of both the trace and the substrate, the latter e.g. for the analysis step of ACE-V see Section 3.4.1). Annotating the digitised trace and substrate (including trace location on the substrate's surface) is provided by methods for use in digitised forensics SF that we define as a set of methods for data presentation and annotation of evidence PA. The following Table 49 (based on Table 40) summarises the main characteristics of the set of methods for data presentation and annotation of evidence PA and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	**Description according to [KDV15]**	**Context to the forensic process in digitised forensics SF**
Methods for presentation and annotation of evidence (PA)	Contains methods that support the manual work of forensic experts by highlighting important trace features	Annotation of the trace (and substrate) image with features and other properties

Table 49: Characteristics of the sets of methods for data presentation and annotation of evidence PA in digitised forensics SF based on [KDV15, p. 89]

Even more vital than for digital forensics, in digitised forensics severe consequences arise from the examination results as they can positively tie a person to a crime scene, potentially leading to a substantial sentence. Great care has to be taken to ensure an examination that is comprehensible, that the chain of custody and the chain of evidence are maintained (see Section 2.4). In a digitised environment, several approaches to maintain a chain of custody are known, e.g. the forensic container [KVL11, p. 266] or the FIDEX format [For12, p. 1]. The all employ methods that we define as a set of methods for chain of custody maintenance CC. The following Table 50 based on Table 40 summarises the main characteristics of the set of methods for chain of custody maintenance CC and adds context for the forensic process in digital forensics SF.

Sets of methods for the forensic process in digitised forensics	Description according to [KDV15]	Context to the forensic process in digitised forensics SF
Methods for chain of custody maintenance (CC)	Contains methods that support ensuring the security aspects of integrity and authenticity as well as the process accompanying documentation, e.g. used sensor or filter settings	Maintenance of chain of custody and chain of evidence with dedicated forensic file formats and containers

Table 50: Characteristics of the sets of methods for chain of custody mainenance CC in digitised forensics SF based on [KDV15, p. 89]

In the following we take a step back and reflect on what we have achieved regarding the sets of methods and in the Section 4.5 about the findings from Section 4.

4.3.3 Reflection and parting thoughts

For both digital forensics DF in Section 4.3.1 and for digitised forensics SF in Section 4.3.2 we applied the first part of M3 as part of our methodology (see Section 1.2).

> M3:
> We use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, p. 4].

We looked at means to describe forensic tools and methods as a set of transfer functions operating on the forensic data types for digital forensics DF and digitised forensics SF as contributions C3 (including C3.1 and C3.2) as presented in Section 1.3.

> C3:
> Establishment of a hierarchical mutual exclusive categorisation of methods for the forensic process:
>
> C3.1:
> Distinction into 6 distinct classes based on the likeliness of availability for digital forensics for selected use cases
>
> C3.2:
> Distinction into 10 classes based on the pipeline of the biometric process for a use case

By encapsulating sets of methods into residual class structures alongside the likeliness of presence for digital forensics and alongside their places inside the signal processing and pattern recognition pipeline for digitised forensics we found a means to abstractly describe their essence without references to specific products and manufacturers.

In the following sections we look into the encapsulation of steps taken throughout the forensic examinations to extend the comprehensibility of our description of the forensic process.

4.4 Sets of examination steps for the forensic process

In informal starting point and attempt to categorise sets of examination steps is our work in [KHD09, p. 1], which we extend in the following.
In [InR01, p. 79] a forensic event reconstruction is described as an ordering of associations in space and time, with associations being the inference of contact between a source of evidence and a target as described Section 2.2. We will use this as a simile when we describe the set of examination steps for the forensic process as a means for the ordering of forensic data types for digital forensics DT and for digitised forensics DD and sets of methods for the forensic process for digital forensics DF and for digitised forensics SF in space in and time, furthering comprehensibility of the forensic process. Rendering the examination process comprehensible is not only a demand from the forensic science itself, it also allows to look for alternatives (e.g. in case of unavailability of certain forensic tools and methods).
We recall the research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

To order the forensic data types and the sets of methods for the forensic process in space and time, we define sets of examinations steps based on [BeC05, p. 149] as follows:

> Sets of examination steps:
> Sets of examination steps are discrete steps in the forensic process and usually a function of time, suggesting a necessarily sequential and sometimes iterative approach (based on [BeC05, p. 149]). They encapsulate related actions and are founded on a pipelined description of the forensic examination process. Sets of examination steps are mutually exclusive, thus forming a residual class structure to be unambiguous.

They represent the practitioner's view according to [PBM08, p. 114] on the grounds that they are process oriented and derived from existing procedures from traditional forensics. The horizontal arrangement of categories in the Forensic Examination Taxonomy [AKD09, p. 64] as outlined in Section 2.5.2 of this thesis reflects the knowledge gathered when following the examination steps for digital forensics DF. These are described in the following Section 4.4.1 and as the contents of the steps differ for digitised forensics SF, they are described in Section 4.4.2.

4.4.1 Sets of examination steps for digital forensics

The sets of examination steps described in the following are the result of a detailed literature study and a subsequent comparison based on new literature performed in Section 3.1. They are first proposed to a German speaking audience in [KHA+09, p. 478] and to an English speaking audience in [KHD09, p.1] and summarised and refined in [KDV15, p.88]. The Table 51 summarises the identified sets of examination steps together with a short content description based on [KDV15, p. 88]. Note that we divert from the often discrete usage of analysis and examination, as pointed out by the reviewer Eoghan Casey, and also define the terms examination and investigation in a manner that the whole of the process is referred to as the examination (see also [Pol08, p.18]) and the investigation is a discrete step within this process (see also Section 2.1.1).

As we point out in Section 4.5, the qualitative estimation of loss, error and uncertainty (see Section 4.1) requires the forensic examiners to test and evaluate their sets of methods for the forensic process extensively to have a notion of a ground truth as to which data types form the data of the investigation target D_{IT} and which data of the investigation results D_{IR} are to be expected. This is why we extend the strategic preparation for forensic examiners compared to [KDV15, p. 88].

Sets of examination steps	Content Description based on [KDV15]
Strategic preparation (SP_{DF})	Includes measures taken by the operator of an IT system and by the forensic examiners in order to support a forensic investigation prior to an incident
Operational preparation (OP_{DF})	Includes measures of preparation for a forensic investigation after the detection of a suspected incident
Data gathering (DG_{DF})	Includes measures to acquire and secure digital evidence
Data investigation (DI_{DF})	Includes measures to evaluate and extract data for further investigation
Data analysis (DA_{DF})	Includes measures for detailed analysis and correlation between digital evidence from various sources
Documentation (DO_{DF})	Includes measures for the detailed documentation of the proceedings, also for the ”transformation” into a different form of description for the report of the incident

Table 51: Sets of examination steps for digital forensics based on [KDV15, p. 88], note that we divert from the typical discrete usage

If the necessity arises, in theory, those sets of examination steps can be extended. In the following, those sets are described in more detail.

4.4.1.1 Strategic preparation for digital forensics (SP_{DF})

The strategic preparation for digital forensics (SP_{DF}) includes all preparation procedures taken ahead of the actual occurrence of a specific incident. We take the ideas for the pre-incident preparation from [FrS07, pp. 29-30] (see Section 3.1.6) into account and include them in this examination step. Further, we integrate the idea from [CaS03, p. 10] where each electronic device is to be treated as a separate crime scene. We extend this idea even further in Section 6 where it is shown that even desktop IT systems such as workstations, servers etc. are composite devices and that their constituent components could be considered potential crime scenes of their own. The following Table 52 based on Table 51 extends the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	Description according to [KVD15]	Context to the digital forensic process
Strategic preparation SP_{DF}	Includes measures taken by the operator of an IT system and by the forensic examiners in order to support a forensic investigation prior to an incident	Documentation and extension of knowledge of IT systems specifics, tool testing for forensic data types and sets of methods determination for error loss and uncertainty estimation, setup of logging capabilities, performance of system landscape analysis, data protection considerations

Table 52: Description of the examination step of strategic preparation SP_{DF} for digital forensics DF based on [KDV15, p. 88]

In the following, this context to the digital forensic process is outlined. We argue that the knowledge regarding specific IT systems and technology by and large at a given time t_i has to be recorded and taken into consideration. We identify three different perspectives as yet:

- the knowledge of the components manufacturer at t_i,
- the knowledge of the attacker at t_i and
- the knowledge of the forensic examiner at t_i.

Trivially, the manufacturers of hard- and software have to have certain knowledge to include options of system design and configuration, which can be used by attackers and by forensic examin-

ers alike. For instance, if manufacturers of hardware would not know that they can update software in place by using Flash/EEPROM as storage for their firmware and thus would use mask-programmed ROM or UV-erasable ROMs (EPROMs), the malicious software-only firmware alteration would be impossible.

Also the knowledge of the attacker obviously needs to include the attack vectors for a successful component compromisation. So if there is no knowledge e.g. about debug interfaces or of the possibility to alter firmware permanently, attackers cannot compromise components and finally IT systems in this way.

As needed for a successful event reconstruction in forensics the knowledge of the attacker forensic examiner also has to include potential attack vectors. This is also important when new facts about a specific case surface even years after its first examination come to light, potentially calling for a re-examination if new facts regarding the knowledge (see also [Ada19] for a discussion on how the growing knowledge can de-mystify event reconstruction). Recording the knowledge of all parties can greatly support a re-examination, as the assumptions can be revisited with new knowledge.

Generally, during strategic preparation in digital forensics SP_{DF}, in reference to the data modelling in Section 4.1 only data containing case-specific, forensically relevant data D_{IFC} are positively ruled out as SP_{DF} is executed ahead of a given case.

A key part of strategic preparation is the definition of forensic data types (see Section 4.2.1) and sets of methods for the forensic process (see Section 4.3.1). The forensic data types and sets of methods for the forensic process described in this thesis are shown to fit the examples of Sections 5 to 6. Other targets may need an addition of data types or alternative definitions (see e.g. [AHK+19, p. 135]). As stated in [RGM+06, p. 39], computer forensic examiners need a *repertoire of tools* and, just as important, a *repertoire of examination and investigative approaches*. This becomes highly important for the estimation of loss, error and uncertainty as introduced in Section 4.5. Programmes like the computer forensic tool testing programme [Gray12, p. 1] of the National Institute for Standards and Technology (NIST) need to be used to determine which data types from digital forensics DF are suited data of the investigation target D_{IT} of a given method of the forensic process and which data types of the data of the investigation result D_{IR} are to be expected (see also appendix Section 10.6).

One can distinguish between measures of SP_{DF} that address forensic capabilities of the forensic investigators and their tools *in general* (i.e. independent of a specific system) from measures of SP_{DF} that are taken *on a specific system* to enhance the possibilities to find traces of incidents (e.g. extra and secure logging mechanisms, stealthy access to network connections). In [CaS03, p. 7] this differentiation is provided using the terms *Operational Readiness Phase* for the former and *Infrastructure Readiness Phase* for the latter. By not limiting ourselves to technical measures but adding organisational and personnel measures, this definition is broader than *forensic readiness* (see e.g. [Tan01a, p. 1]) but shares its objectives. However, as remarked by the reviewer Eoghan Casey, forensic readiness also implies some important elements such as asset inventory, policy setting, response planning and can help mitigating loss, error, uncertainty but also alteration [CaS20, p. 2-6].

Also, the *set up of the logging capabilities* and thus, what to log on an IT system consisting potentially of a multitude of (networked) components (including the network infrastructure elements), occurs during strategic preparation. This is based on the comprehensive view of digital forensics that includes the operator of an IT system as a key factor for the degree of success in an examination based on the strategic preparation performed (see Section 2.1.1), which is also included in the models of [KCG+12], [FrS07] and [CaS03] as reviewed in Section 3.1. Logging can include the establishment of SNMP capabilities (see Section 4.2.1.3.2). As stated in [PBK+07, p. 3], the challenge for this setup is to balance the usefulness of logged data against the practicality of performing both logging and auditing. The approach presented in [PBK+07, p. 4] focuses on determining which data is necessary to understand past events and proposes strategies to log such rel-

evant data. This is clearly aimed at the operators of IT systems, which have the ability to alter the logging process ahead of an incident. Such an approach would not work for examiners (e.g. from the police force), which are called to examine an incident that has already happened. Determining the precise means to optimise the type and amount of the logged data for specific potential attacks is beyond the scope of this thesis, [PBK+07] is suggested for further reading on this subject.

The *system landscape analysis* (including the planning of the IT system) and the comprehensive documentation of the hardware and software components together with their networking is done during strategic preparation, i.e. ahead of (suspected) incidents. It includes the enumeration and configuration as an inventory of hard- and software components forming immediate the IT system of the operator and as much information as possible regarding external components and services. By placing this step into strategic preparation (ideally) and thus into the custody of the operator of an IT system, we pick up the suggestions of a digital evidence map suggested in [Cas11, pp. 209-210]. This map, according to [Cas11, pp. 209-210, p. 643], should indicate where evidence is located on a network (including e-mail, log files, etc.). It is a graphical description of the network and where potential sources of evidence are located [Cas11, p. 643].

We extend the digital evidence map (and call it system landscape analysis henceforth) to all data streams (see Section 3.3) by adding:

- hard-/software inventory (including sets of methods for the forensic process),
- placement of hard- and software sensors in aid of intrusion/anomaly detection,
- identification of protection requirements.

Figure 36 shows a simplified exemplary system landscape analysis for a small client/server IT system (based on [BSI20, p. 83]).

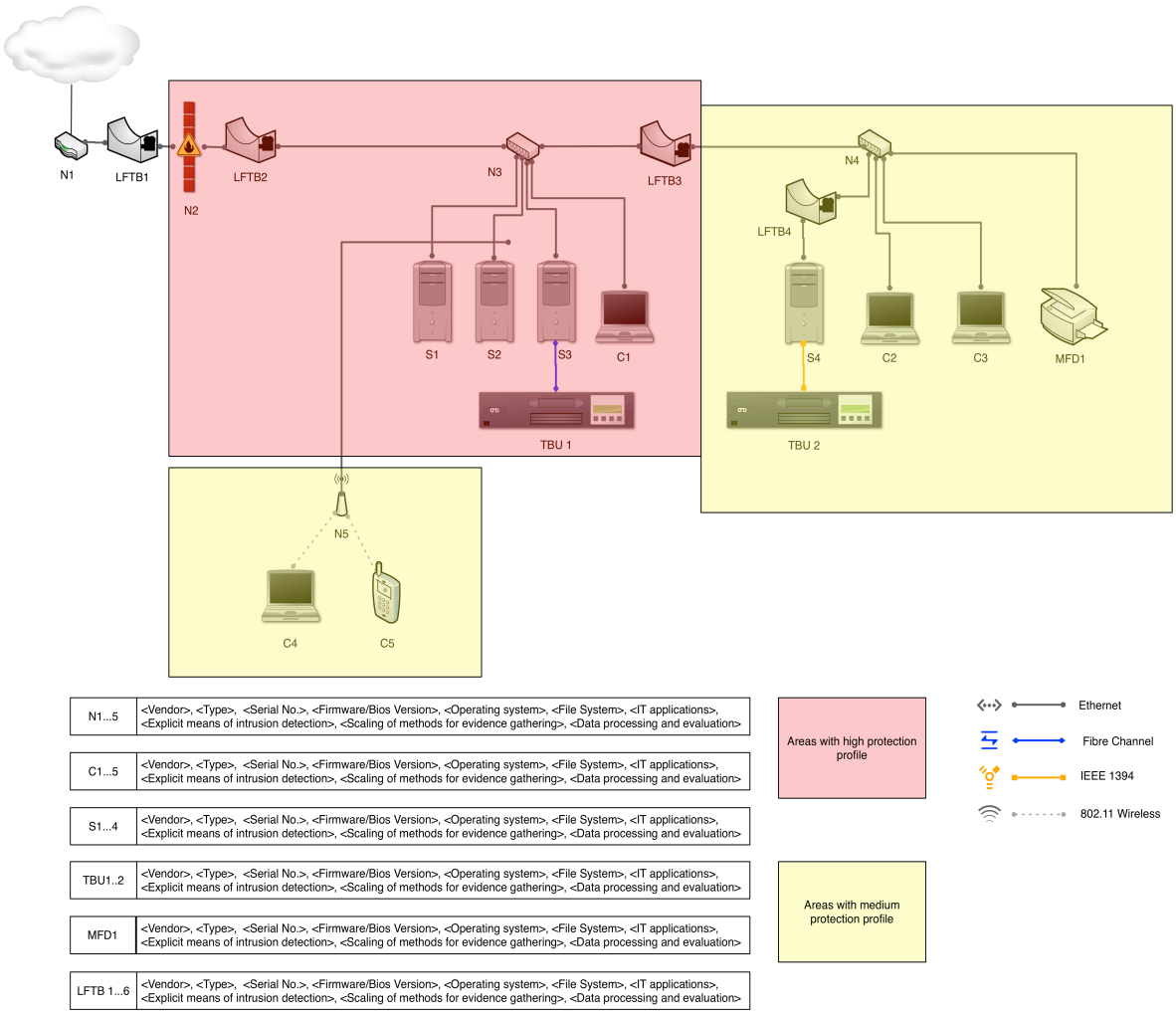

Figure 36: Simplified exemplary system landscape analysis based on [Cas11, p. 644] and [KHA+10, p. 100] comprising a HW/SW inventory including Multi Function Devices (MFD) and Tape Backup Units (TBU), placement of HW/SW sensors including Linux Transparent Forensic Bridge (LFTB) and identification of protection requirements (derived from [BSI20, p. 83]).

We propose to include the sets of methods for the forensic process for digital forensics (Section 4.3.1) to be listed in the inventory as they determine, which potential forensically data will be available for gathering and processing in the event of a forensic examination. We add relevant locations for the placement of Linux Transparent Forensic Bridges (LFTB, see [KHA+10, p. 100]) as a means to acquire network traffic data in a forensic sound manner (see Section 3.3.1.5). The LFTB is one instance of a hw/sw device that acts as a sensor for network traffic. In this system landscape analysis, additionally, protection requirements for the information processed, stored and communicated are depicted, which are drawn from the security aspects of confidentiality, integrity, availability, authenticity and non-repudiation. This is based on the proceedings outlined in the BSI-Standard 200-2 [BSI20, p. 83]. The more detailed the analysis is, the more comprehensive and meaningful the estimation of loss, error and uncertainty gets. The system landscape analysis needs constant updating once des system is in use as its hardware and software components are likely (and often required) to receive updates/upgrades. The planning includes the placement of hard- and software sensors (e.g. for host-based and/or network based intrusion detection systems). Such systems, besides their main intended function of preventing or containing incidents, also log potentially forensically relevant data. Further, the setup of a time base and maintenance of proper clock synchronisation and detection of deviations on all involved IT systems to implement accurate time stamps is vital for digital forensics. The result of all of the above

measures is the definition of a system landscape, which is also important for an initial boundary estimation in case of an incident.

Strategic preparation and thus forensic readiness becomes increasingly important for incident investigation outside the law enforcement environment. Large organisations must establish and maintain forensic capabilities whilst adhering to the full extend of data protection legislation. In [Dil18, pp. 118-119] the term *enterprise forensic readiness* is used to describe this capability. In [KDV15, p. 87] the idea of forensic readiness is broadened to not only focus on forensic investigation processes using existing tools but also to cover design processes for new methods for the forensic process and ultimately to the design process for systems design (HW/SW), i.e. the investigation target. Strategic preparation can impact privacy, which is why data protection measures (see Section 2.7) need to be implemented in this step as well.

Generally, both system administrators preparing for potential attack as well as examiners need to *prepare storage capacity* for forensic duplicates and forensic images see Section 3.3.1.5. They need to make sure that the media is properly sanitised (see Section 3.3.1.6) by forensic sound deletion to avoid cross-contamination of evidence from other cases or other non-case related media content, which could lead to errors and thus evidence dynamics (see also [Cas11, p. 27]).

All actions taken should be comprehensively documented (see also Section 4.4.1.6).

4.4.1.2 Operational preparation for digital forensics (OP_{DF})

The operational preparation for digital forensics (OP_{DF}) is defined in [KHD15, p. 88] to include all preparation procedures taken after of the actual occurrence of a specific incident. Those procedures by definition do not alter any data on the targeted system. The following Table 53 based on Table 51 contains the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	**Description according to [KVD15]**	**Context to the digital forensic process**
Operational preparation OP_{DF}	Includes measures of preparation for a forensic investigation after the detection of a suspected incident	Only external data containing case-specific, forensically relevant data D_{IFC}, Selection and justification of sets of methods for the forensic process for the gain of case-specific, forensically relevant data D_{IFC} against the structural impact, development of search strategies and objectives of the examination and determination of depth of association, initialisation of the chain of custody, digitalisation of external information by the examiner, updating of the system landscape description, localisation and level of detail selection of potential data streams DS based on system landscape description,

Table 53: Description of the examination step of operational preparation OP_{DF} for digital forensics DF based on [KDV15, p. 88]

In the following, this context to the digital forensic process is outlined. We incorporate the formulation of an examination plan for data collection and analysis as proposed by [BeC05, p. 150] that is summarised in Section 3.1.9.

Generally, during operational preparation in digital forensics OP_{DF}, only external data are to be expected. This observation is based on the grounds that no data from the data stream sources have been gathered, yet. Those external data created during the examination process (e.g. case number etc.) need to be digitised by the examiner to be labelled clearly in the documentation step, see Section 4.4.2.6

A very stringent criterion of the measures chosen in the following steps is the estimation of their impact on the investigated system, i.e. which forensic data types DT (Section 4.2.1) as data of the investigation target D_{IT} from what data stream DS could contain the most case-specific, forensically relevant data D_{IFC} (Section 4.1) whilst weighing that with the structural impact (Section 3.3.1.3) of the chosen set of methods for the forensic process. Although often case-specific requirements exist (e.g. in live forensic investigations often in less-than-ideal circumstances), generally the data collection procedures (see Section 4.4.2.3) with the least and best predictable impact on the data within the investigated system are to be preferred. This includes the weighting process if a system is networked and running at the time of the suspected incident (see Section 3.1.4). As stated in [KH02, p. 5-6], if the system is running, a case-specific decision has to be reached whether the system should be shut down or left running. The latter obviously denies any investigation of items of high volatility (see also Section 3.3.1.1), which includes the main memory stream. Further it could deny access to information contained in data from mass storage devices if encryption techniques are used on the devices and the keys are unknown. On the other hand, the shutdown could be the only sensible solution to contain the consequences if the incident is ongoing and the running processes render data in the system unintelligible (e.g. in ransomware attacks). Especially for large server systems even the type of shutdown is important (orderly shutdown vs. power cord/main battery removal). But, and this goes further than the discussion in [KH02, p. 5-6], also the potential disconnection from the network or the alteration of network settings is a weighted process since malware exists, which is able diagnose such alterations and change both the behaviour of the malware or the system under examination. In any case, each of these case-dependent decisions need to be well justified and recorded in the process accompanying documentation as part of the documentation step DO_{DF} (see Section 4.4.1.6).

As stated in [RGM+06, p. 37], the value of planning, i.e. the operational preparation in this thesis cannot be over-emphasised. For instance, reliable information on search terms (often very case-specific), contacts, types of activities, hard- and software used etc. in advance of the search can allow the conductor of a forensic examination to develop at least some search strategies before the data gathering and subsequent examination and analysis. In time-critical situations, such as a triage-based examination (see [RGM+06, p. 37] and its review in Section 3.1.8), every minute saved is potentially another minute available to conduct the search on site itself. Also the formulation of an approach strategy at the start of an examination and its continued refinement as introduced by [BeC05, pp. 153-156] and reviewed in Section 3.1.9 are to be considered. Further, the objectives of the examination have to be stated (e.g. criminal proceedings but also operational troubleshooting and other purposes, see Section 2.3), which have profound influences in the depth of the association (see Section 2.3). This necessitates a defendable and justifiable decision about the depths of association to be taken from operational preparation onwards (see Section 4.4.1.2) as to how deep the forensic examination has to look. The objectives of the examinations are limited not only by resources (time, money, expertise) but also by a number of legal requirements, not least of which is data protection (see Section 2.7).

An important part of the operational preparation is the initialisation of the case-specific process accompanying documentation (see Section 4.4.2.6). This includes the initialisation of 'absolute' clocks (see Section 3.3.1.4) and the provision of sufficient external storage capacity and potentially the initialisation of forensic evidence storage structures therein (e.g. a digital evidence bag [Tur06, pp. 59-64] or the AFF4 format [CGS09, pp. 57-68]). Necessary cryptographic keys need to be initialised/managed together with further means to adhere to the demands from the Cyber Forensic Assurance model [Dar10, p. 62] (see Section 3.1.1), notably the maintenance of the components of integrity and authenticity. This initialisation process implements the start of the chain of custody for (non-tangible) digital data (see also Section 2.4 and Section 3.1.4). Additionally, all tangible, physical devices potentially containing data from which information about an incident can be extracted need to be included into the chain of custody. This entails gaining information describe those devices generally and uniquely (e.g. model, HW/SW revision, serial number) that can be retrieved by visual inspection. Note: Although this information is sometimes

electronically extractable and should be extracted, this is part of data acquisition (see Section 4.4.1.3) and can potentially alter data contained in these devices.

Often, some information (e.g. model serial number, HW/SW revision) is also printed non-volatile on the devices (e.g. using labels). That information can be used to validate the data gathered in the next step (primarily hardware data DT_2, see Section 4.2.1). However, it needs to be stated that from the viewpoint of the Data-Centric Examination Approach (DCEA) this data is not part of the digital data of any of the data streams and thus constitutes *external data*, which needs to be digitised by the examiner to be labelled clearly in the documentation step (see Section 4.4.1.6). The need for validation is justified in Section 6. However, the labelling of then refers to factory conditions (e.g. BIOS revision information) and can be altered by the user as part of normal usage and maintenance. The challenge to differentiate between such alterations and malicious alterations as shown in Section 6 remains.

If time permits, an update of the description of the current system landscape analysis could prove to be valuable, especially for initial boundary estimation of the (suspected) incident.

From the system landscape analysis from strategic preparation (SP_{DF}) or in less-than-ideal circumstances via an ad-hoc system landscape analysis, *potential data sources of all data streams* (mass storage, main memory, network, Section 4.2.1) are located together with potential data gathering methods that minimise the impact to the investigated system and maximise available data.

At this step (at the latest), the *level of detail selection* (ideally based on the system landscape analysis performed and updated during strategic preparation, see Section 4.4.1.1) needs to be established, which defines the granularity of the findings. Consider a networked client/server based desktop IT system, for example. The level of detail is expected to be less, if mass storage systems (see Section 6) are only treated as block-level dumb storage system instead of processor-controlled components. Some (potential crucial) information might be missed, if e.g. the mass storage systems themselves are not treated as an IT system in its own right. However, economical and/or time resource constraints might necessitate a more coarse level of detail selection (as is, indeed, the case in the scenarios of this thesis apart from the one described Section 6). This coarse level could be the assumption, that desktop IT devices are atomic units and the inner structure resulting from the use of processor-controlled components is disregarded (see e.g. Section 5). Of course, the level of detail selection has a profound influence on error, uncertainty and loss (see Section 4.1) and thus need to be recorded and justified as part of process accompanying documentation (see Section 4.4.1.6).

Thanks to the remark of the reviewer Eoghan Casey we recognise that some process models place this step, which contains actions from the survey of the incident scene (see [Cas11, p. 240-245]) as prior to the examination whilst we include this in the modelling of the examination on the grounds that the process accompanying documentation receives data from this step.

All actions taken should be comprehensively documented (see also Section 4.4.1.6).

4.4.1.3 Data Gathering for digital forensics (DG_{DF})

In the examination step of data gathering DG_{DF}, the data is acquired (gathered) as a basis for all following examination steps. It needs to be secured primarily with regards to the security aspects of authenticity and integrity but often also confidentiality, especially if (as is to be expected) person-related data is to be gathered. This step is mentioned using other names in existing models (see Section 3.1) e.g. Acquisition [New07, pp. 5-7] or Preservation [CaS03, p. 4]). The following Table 54 based on Table 51 contains the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	Description according to [KVD15]	Context to the digital forensic process
Data Gathering DG_{DF}	Includes measures to acquire and secure digital evidence	Initial collection of data as the basis for all subsequent case-specific examination steps for the extraction of data containing case-specific, forensically relevant data DIFC using all sets of methods for the forensic process operating on all data types for digital forensics DF (ideally on raw data DT_1), potentially only gathering a subset of available data using triage techniques, (ideally atomically) maintenance of integrity, authenticity and confidentiality during the gathering step, potential update of the system landscape analysis

Table 54: Description of the examination step of data gathering DG_{DF} for digital forensics DF based on [KDV15, p. 88]

In the following, this context to the digital forensic process is outlined. Based on the decisions taken during operational preparation OP_{DF}, in the examination step of data gathering DG_{DF} data is acquired. It forms the initial and thus most basic step, which is performed on-site involving the target IT system or its network communication. The resulting data represent forensic duplicates (see Section 3.3.1.5) used for further analysis. Any loss, error or uncertainty (see Section 4.1) is likely to have grave consequences especially on data with a high volatility (see Section 3.3.1.1). All data streams DS (see Section 4.2.1) together form the data source for this acquisition, potentially containing case-specific, forensically relevant data D_{IFC} (Section 4.1). This gathering process can use all methods from the set of methods for the forensic process in digital forensics DF (Section 4.3.1) and operates on the data of types DT of the data streams as data of the investigation target D_{IT} and return data of the investigation result D_{IR}. Ideally in the 'rawest' data type, that is raw data DT_1 should be collected to have access to all the other data types contained therein (see Section 4.2.1).

Data gathering can be performed post-mortem or during live forensics (see Section 3.3.1.2), the latter posing the problem of time and resources and emphasising the need for a comprehensive operational preparation OP. The resource problems stem from the fact that modern mass storage devices have grown substantially faster with regards to capacity compared to data interface speeds (see e.g. [Sch20]). One way to address this problem is the introduction of the triage to focus on parts that are deemed to contain case-specific, forensically relevant data D_{IFC} (see e.g. Section 3.1.8 and [RGM+06, pp. 32-33]) whilst disregarding parts that are deemed to only contain data containing forensically relevant information D_{IF} or plain data containing information D_I.

To maintain the IT security aspects, methods for the forensic process for data gathering need to augment the gathered data with data generated from its intrinsic content for maintenance of integrity and authenticity purposes to support the step of process accompanying documentation (see Section 4.4.1.6). Ideally, those mechanisms are atomically integrated into the method used for data gathering as opposed to added afterwards, which can raise questions.

During the step of data gathering DG_{DF}, an update of a system landscape analysis (see Section 4.4.1.1) can be triggered, if concealed or hidden devices containing data are discovered.

4.4.1.4 Data Investigation for digital forensics (DI_{DF})

In the examination step of data investigation for digital forensics DI_{DF}, measures are taken to evaluate and extract data (based on the data gathering, Section 4.4.1.3). The following Table 55 based on Table 51 contains the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	Description according to [KVD15]	Context to the digital forensic process
Data Investigation DI_{DF}	Includes measures to evaluate and extract data for further investigation	Data reduction to limit or exclude data containing information D_I and data containing forensically relevant information D_{IF} leaving only data with case-specific, forensically relevant information D_{IFC}, performed on forensic duplicates fostering repeatability with other methods also affecting loss, error and uncertainty, often highly automated methods confined to the stream of the forensic duplicate, potential update of the system landscape analysis for further examinations

Table 55: Description of the examination step of data investigation DI_{DF} for digital forensics DF based on [KDV15, p. 88]

In the following, this context to the digital forensic process is outlined. The examination step of data investigation DI_{DF} intends to remove data that contains information D_I and data containing forensically relevant data D_{IF} in an effort to only retain data containing case-specific, forensically relevant data D_{IFC}. Often this amounts to a significant reduction of data sizes compared to the data sizes after the data gathering. The examination step of DI_{DF} ideally (see Section 4.4.1.3) starts with the data type DF of raw data DT_1 (Section 4.2.1) from any data stream DS and can then branched off or re-iterated/refined in order to progress towards the limitation to only containing case-specific, forensically relevant data D_{IFC}. In theory all methods of the forensic process (Section 4.3.1) can be used with all data types DT of digital forensics DF acting as data of the investigation target D_{IT} and as data for the investigation result D_{IR}. The inclusion of raw data DT_1 is relevant due to the iterative nature of the data investigation step. One important characteristic of data of the investigation result D_{IR} is derived from gathered data and thus not available in the initial system states during the data gathering.

The data investigation step DI_{DF} often is performed off-site using forensic duplicates of the gathered data (see Section 3.3.1.5). This way, alternative investigation approaches can be tested, e.g. in an effort to reduce loss, error and uncertainty (see Section 4.1) and to verify the results by independent experts. However, the step data investigation DI_{DF} can also a part of live forensics (see also Section 3.3.1.2) for selected data items, potentially triggering a full scale forensic investigation afterwards.

The step of data investigation DI_{DF} is often highly automated, e.g. the resurrection of deleted data in file systems and the meta data extraction (Section 4.2.1.1.2), the extraction of process data from a memory dump (Section 4.2.1.2.2) or the re-assembly of TCP/IP data streams in networks (Section 4.2.1.3.2). Also, the process of data investigation remains in the confines of the same data stream as the forensic duplicate as data of the investigation target D_{IT} originated from.

Generally, the data investigation DI_{DF} step often follows a searching/extraction loop [RYG05, p. 3], whereby a seed information search leads to data filtering and decision making and subsequent information extraction. This in turn leads to a refinement of the query, which results in a new seed information search.

The data investigation DI_{DF} is also a part of live forensics (see also Section 3.3.1.2) for selected data items (e.g. during a triage, see Section) potentially triggering a full-scale forensic examination afterwards.

During the step of data investigation DI_{DF}, an update of a system landscape analysis (see Section 4.4.1.1) can be triggered, if new data sources containing data are discovered, potentially triggering a new data gathering step.

4.4.1.5 Data Analysis for digital forensics (DA_{DF})

In the examination step of data analysis for digital forensics DA_{DF}, measures for detailed analysis and correlation between digital evidence from various sources are taken (based on the data investigation DI_{DF}, Section 4.4.2.4). The following Table 56 based on Table 51 contains the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	Description according to [KVD15]	Context to the digital forensic process
Data Analysis DA_{DF}	Includes measures for detailed analysis and correlation between digital evidence from various sources	Reduction of data to only contain data containing case-specific, forensically relevant data D_{IFC}, semi-automated intra- and inter-stream correlation of data according to timelines and cause/effect relations, potential update of system landscape analysis and further examinations

Table 56: Description of the examination step of data analysis DA_{DF} for digital forensics DF based on [KDV15, p. 88]

In the following, this context to the digital forensic process is outlined. The aim of the examination step of data analysis for digital forensics DA_{DF} is to further narrow down on the data containing the investigation result D_{IR}, which in theory can contain all data types for digital forensics, from the preceding step of data investigation DI_{DF} and further remove data containing information D_I and data containing forensically relevant information D_{IF}. The goal is to obtain only data containing case-specific, forensically relevant information D_{IFC}. In theory all methods of the forensic process for digital forensics can be used. However, one expects not to gain raw data DT_1 as the data containing the investigation result D_{IR} when applying those methods in the examination step of data analysis DA_{DF}.

This step is less likely to be fully automated, as it requires some human ingenuity. In the step of data analysis, data from other data streams DS (see Section 4.2.1) can be used to correlate events that leave different traces in different data streams (inter-stream). The idea is to substantiate data from all possible sources. Kirk as cited in [InR01, p. 61] states: 'A single piece of evidence is rarely sufficient in itself to establish proof of guilt or innocence.' But also if only one data stream is available, correlation occurs between the different data types (intra-stream), which also reflect events in a data stream differently. As motivated by [InR01, p. 79], the ordering of associations in space and time, with associations being the inference of contact between a source of evidence and a target, has to occur for an examination to be successful. Transferred to the digital world, it can be argued the equivalent is the ordering of the existing data containing case-specific, forensically data D_{IFC} such that it enables event reconstruction. Most prominently this ordering is achieved according to a timeline or a cause/effect relation), which is ever more important if data from different data stream is correlated.

The execution of the data analysis examination step DA_{DF} is also a part of live forensics (see also Section 3.3.1.2) for selected data items (e.g. during a triage, see Section) potentially triggering a full-scale forensic examination afterwards.

During data analysis DA_{DF}, an update of a system landscape analysis (see Section 4.4.1.1) can be triggered, if new data sources containing data are discovered, potentially triggering a new data gathering step.

4.4.1.6 Documentation for digital forensics (DO_{DF})

The documentation step serves two purposes. During the forensic examination and throughout the forensic process, thus, including the step of strategic preparation (see Section 4.4.1.1), it serves as technical documentation facility to (semi-) automatic record every digitally available item as a result of forensic proceedings (*process accompanying documentation*).

After the forensic examination, from the results of this documentation step, the *final report* is compiled. There is likely to be a number of final reports, especially with regards to technical detail, depending on the intended audience.
The following Table 57 based on Table 51 contains the description from [KDV15, p. 88] and adds context to the digital forensic process.

Sets of examination steps	Description according to [KVD15]	Context to the digital forensic process
Documentation (DO_{DF})	Includes measures for the detailed documentation of the proceedings, also for the ”transformation” into a different form of description for the report of the incident	Operate on all data types, data of the investigation result D_{IR} same as data of the investigation target D_{IT} or excerpts from it and/or enriched with meta data from the examination process

Table 57: Description of the examination step of documentation DO_{DF} for digital forensics DF based on [KDV15, p. 88]

Both the process accompanying documentation and the final report as parts of the examination step of documentation DO_{DF} operate on all data types of digital forensics DF as data of the investigation target D_{IT}. They are limited to the data of case-specific, forensically relevant data D_{IFC} and forensically relevant data D_{IF}. Data of the investigation result D_{IR} are the data themselves enriched with meta data regarding the examination process.

All sets of examination steps are sources for the process accompanying documentation as part of the examination step of documentation DO_{DF}. The documentation starts with the initiation of the forensic process. Case independent data is recorded from the strategic preparation SP (see Section 4.4.1.1) and forensically relevant data D_{IF}. The recording of case-specific, forensically relevant data D_{IFC} starts with the initialisation of the chain of custody for digital objects in the examination step of operational preparation OP_{DF} and the implementation of the concept of provenance (see Section 2.4), which is partly addressed by our Data-Centric Examination approach DCEA. The sets of methods employed are of the data presentation and evaluation DPE (see Section 4.3.1) variety. They often implement container structures (see appendix Section 10.5, e.g. Digital Evidence Bags, [Tur06, pp. 59-64]), which typically handle the maintenance of the IT security aspects of integrity, authenticity and confidentiality as well as adding meta data such as insertion time, inserting author, location etc. Due to the overarching nature of the process accompanying documentation it takes *principles* from [BeC05, p. 152] as overarching goals and to define procedures, guidelines and/or methodological approaches overlapping all sets of examination steps (see Section 3.1.9). A very important constituent of process accompanying documentation is the configuration data of the respective methods used (the rule set from [Car03, p. 4]). Also the recording of the level-of-detail selection (in operational preparation, Section 4.4.1.2) is very important. Thanks to the remark of the reviewer Eoghan Casey we point out that in each step a thorough evaluation as outlined in [PCJ+18, p. 7] has to take place, i.e. producing values that can be fed into a decision process. This decision process involves assigning a strength of evidence assignment (see e.g. [Cas19, pp. 1-10]) and needs to be recorded in the process accompanying documentation. Further, the final evaluation, which can also entail event reconstruction, should also manifest itself as a discrete step in itself. However, the latter is out of the scope of this thesis as outlined in Section 1.

For the final report, excerpts of the data (e.g. human readable listings, forensic tool output, images contained in or communicated the target IT system) are used together with meta information that describes the process leading to the results documented. Errors, loss and uncertainty is always to be expected, especially if human experts are involved and humans who design the software both of the target IT system and that of the examiners. Thus, losses, errors and uncertainties must not be covered up but instead explained by the forensic expert instead together with a qualified judgement of their (likely) implications. We believe the Data-Centric Examination Approach DCEA introduced in thus thesis can be great support for those judgements. The main purpose of

the final report is the communication of relevant findings to a variety of audiences (technical, legal personnel, law enforcement, management etc., see [BeC05, p. 151] as described in Section 3.1.9).

4.4.1.7 Flow of the examination process in the context of the examination steps

The examination steps for digital forensics DF describe actions that are executed by the forensic examiner using methods of the forensic process (Section 4.3.1), which operate on forensic data types DF (Section 4.2.1). They are connected to one another by providing data and information for the following step. The following Figure 37 depicts the intended flow of the examination process according to the Data-Centric Examination Process (DCEA).

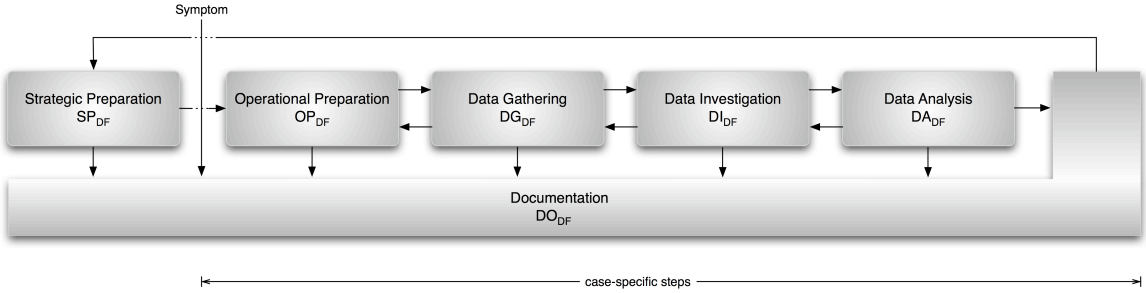

Figure 37: Flow of examination steps in digital forensics with backward steps and the distinct, case-independent step of strategic preparation SP_{DF}, visualised and extended from [KHD09, p. 1]

Inspired by the models from [New07, p. 6], [CaS03, p. 10], [FrS07, p. 21] as discussed in Section 3.1 we allow for backward steps during the examination. This is often especially necessary when correlating findings between the respective data streams (mass storage, main memory and network, see Section 4.2.1). We allow for backward steps all the way back to operational preparation OP_{DF} since we expect some results in some examinations may lead to a broadening of the search to include previously disregarded or unknown data sources (e.g. external USB storage) for which a dedicated operational preparation is just as necessary.

Very important is that each examination step adds to the process accompanying documentation as part of DO_{DF}. Only with a proper and minute recording of all details both the road to the conclusion and the maintenance of the security aspects can a defendable conclusion be achieved.

Notice the distinction from strategic preparation SP_{DF} from all remaining examination steps (in Figure 37 using a dotted section within the arrows) since it describes only case-independent actions. The results from the case-specific examination presented in the final report as a result of DO_{DF} and lessons learned are used as feedback to enhance the strategic preparation SP_{DF} in future. This is also recommended in [KCG+12, p. 3-1], [FrS07, p. 37] and [CaS03, p. 12]. The case-specific examinations are triggered by a symptom, which embodies a recognised system anomaly of some sort.

4.4.2 Sets of examination steps for digitised forensics

In the following, the sets of examination steps are described, based on the initial grouping outlined to a German speaking audience in [KHA+09, p. 478] and to an English speaking audience in [KHD09, p. 1] and applied to digitised forensics in [HKG+11, p. 4] and summarised in [KDV15, p. 89]. Deviating from the process description proposed in [HKG+11, p. 4], the process of step physical acquisition is omitted for this thesis, as we are only concerned with digital data to operate upon. However, the interested reader is directed to [Hil20] for details about the connection between the physical and digital aspects in digitised forensics. Although the physical acquisition is vital for the whole examination, there is no room for it in a data-centric approach as presented in this thesis. However, documenting the physical acquisition can and should take place digitally, resulting in valuable data. The Table 58 summarises the identified sets of examination steps together with a short content description based on [KDV15, p. 89].

Sets of examination steps for digitised forensics SF	Content description based on [KDV15]
Strategic preparation (SP_{SF})	Includes preparation procedures taken ahead of a specific incident
Operational preparation (OP_{SF})	Includes preparation procedures taken after an incident
Data gathering (DG_{SF})	Includes measures to acquire and secure digital evidence
Data investigation (DI_{SF})	Includes procedures for the extraction of data from the trace carrier and the trace
Data analysis (DA_{SF})	Includes procedures for the separation and visualisation of trace data and trace carrier data
Documentation (DO_{SF})	Includes measures for the detailed documentation of the proceedings, also for the "transformation" into a different form of description for the report of the incident

Table 58: Sets of examination steps for digitised forensics based on [KDV15, p. 89]

If the necessity arises, in theory, those sets of examination steps can be extended. In the following, those sets are described in more detail.

4.4.2.1 Strategic preparation for digitised forensics (SP_{SF})

Similar to the step of strategic preparation for digital forensics SP_{DF} (see Section 4.4.1.1) the step of strategic preparation for digitised forensics SP_{SF} is by definition not case-specific and data obtained therein are non case-specific but forensically relevant (see Section 4.1). SP_{SF} is started well in advance of a suspected crime case (see also Section 4.4.2.7). The Table 59 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	Content description based on [KDV15]	Context to the digital forensic process
Strategic Preparation (SP_{SF})	Includes preparation procedures taken ahead of a specific incident	Contains D_{IT} and D_{IR} from D_I and D_{IF} but no D_{IFC} and tests all sets methods for the forensic process for digitised forensics SF, needs system landscape analysis, trains models for pattern recognition/machine learning

Table 59: Description of the examination step of strategic preparation SP_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during strategic preparation in digitised forensics SP_{SF}, in reference to the data modelling in Section 4.1, the sets of methods used therein operate on all data types of digitised forensics SF (see Section 4.2.2) as data of the investigation target D_{IT} and can return all data types as investigation result D_{IR}. The can consist of data containing information D_I or data containing forensically relevant information D_{IF}. Only data containing case-specific, forensically relevant data D_{IFC} is positively ruled out as SP_{SF} is executed ahead of a given case. These are, based on the grounds that no data from the sensor sources have been gathered yet, external data (i.e. which needs to be digitised by the examiner to be labelled clearly in the documentation step, see Section 4.4.2.6).

The examination step of strategic preparation in digitised forensics SP_{SF} is concerned with the installation and facilitation of fixed sensors and processing equipment in a forensic laboratory and the installation and facilitation of mobile sensors and supporting equipment at the crime scene itself and the benchmarking of the aforementioned items using benchmarking (see also appendix Section 10.6). And since the data processing is proposed to be primarily in the digital domain, this also calls for a planning, documentation and maintenance of the IT landscape (incl. network to-

pology), as described in Section 4.4.1.1. Also the provision of forensically cleaned digital storage for the data gathered during the strategic preparation SP_{DF} and for use in case-specific proceedings (the examination step of operational preparation and onwards) has to be executed here e.g. due to time constraints (see Section 3.3.1.6). A strict isolation between locally handled trace data and data used for comparison (e.g. trace data bases such as the ones used in the automated fingerprint identification system AFIS [YLJ15, p. 67] or the Integrated Ballistics Identification System IBIS [NeR06, p. 18]) need to be assured, especially on the grounds of person-related data protection (e.g. fingerprints, natural fibre structures such as hair etc.) Such material is by definition highly person-related data and must not leave the crime lab by accident. In this context, the forensic laboratory operators are IT system operators (Section 4.4.1.1); only for them it is mandatory to perform system landscape analysis and the connected steps.

Aside from IT security and data protection requirements, also the functional aspects of the tools used (implementing the transfer functions bundled into sets of methods for the forensic process from Section 4.3) needs to be tested. The testing becomes highly important for the estimation of loss, error and uncertainty as introduced in Section 4.5. Tool testing needs to be used to determine which data types from digitised forensics SF are suited data of the investigation target D_{IT} of a given method of the forensic process and which data types of the data of the investigation result D_{IR} are to be expected. Thanks to a remark from the reviewer Jana Dittmann we state that also during strategic preparation the parameter data DD_4 used throughout the whole signal and pattern recognition pipeline (see Section 4.2.2) are optimised in a non-case specific fashion (e.g. by building scan parameters for classes of surfaces with similar optical properties). By definition, only data containing information D_I and data containing forensically relevant information D_{IF} are processed and returned. Further, guidelines established for best performance (e.g. on specific surfaces acting as potential trace carriers and environmental condition). We describe this process as benchmarking and its suggested properties in [HML+11, pp. 1-6] and extend those in [KLD+11, pp. 78810G-78810G-15] and [KHD12, pp. 1504-1508].

Also for all approaches involving pattern recognition/machine learning the models used for the examination of a specific case need to be generated/trained in strategic preparation for digitised forensics SP_{DF} (see also Section 4.3.2).

4.4.2.2 Operational preparation for digitised forensics (OP_{SF})

After a forensic examination has been triggered and before any tool use for the gathering of traces, the examination step of operational examination for digitised forensic OP_{SF} is executed. The following Table 60 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	**Content description based on [KDV15]**	**Context to the digital forensic process**
Operational preparation (OP_{SF})	Includes preparation procedures taken after an incident	Only external data forming data containing case-specific, forensically relevant information D_{IFC}, identification of sensors and methods avoiding trace concurrency, initialisation of chain of custody for digital data and forensic storage formats for integrity, authenticity and documentation

Table 60: Description of the examination step of operational preparation OP_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during operational preparation in digitised forensics OP_{SF}, in reference to the data modelling in Section 4.1 methods of the forensic process SF operate only on *external data* (i.e. data created during the examination process such as case number, description etc.) which needs to be digitised by the examiner to be labelled clearly in the documentation step, see Section 4.4.2.6).

This observation is based on the grounds that no data from the sensor sources have been gathered yet.

During the operational preparation appropriate sensors and methods are identified [HKG+11, p. 4]. Also, a course of action is devised (including both crime scene and forensic laboratory actions). Potential trace concurrency problems need to be detected and avoided (see Section 3.4).
Somewhat analogous to the operational preparation in digital forensics OP_{DF} (Section 4.4.1.2), in the step of operational preparation in digitised forensics OP_{SF}, the case-specific process-accompanying documentation is initiated (Section 4.4.2.6), starting the chain of custody (see Section 2.4). Further, the implementation of the concept of provenance (see Section 2.4) starts, which is partly addressed by our Data-Centric Examination approach DCEA. Note: in this thesis we are only concerned with digital data, similar steps need to be taken for physical objects as well.

Further, means to ensure the IT security aspects of integrity and authenticity for all digital data about to be gathered, investigated and analysed, need to be initiated. Examples are the forensic container from [KVL11, p. 266] or the FIDEX data structure [For12, p. 1].

4.4.2.3 Data Gathering for digitised forensics (DG_{SF})

During the examination step of data gathering for digitised forensics DG_{SF}, data from is acquired using the sensors forming a data stream, respectively. After gathering, the data is stored in forensic data structures to secure the evidence and thus maintain the security aspects of integrity and authenticity. The following Table 61 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	**Content description based on [KDV15]**	**Context to the digital forensic process**
Data gathering (DG_{SF})	Includes measures to acquire and secure digital evidence	D_I or D_{IT} or D_{IFC} as D_{IR} with no digital input data, creation of a forensic duplicate, profile/survey scans, coarse scans and fine scans using the selected sensors, securing the digital data in the chain of custody for digital data

Table 61: Description of the examination step of data gathering DG_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during data gathering in digitised forensics DG_{SF}, in reference to the data modelling in Section 4.1 methods of the forensic process SF operate on data containing case-specific, forensically relevant data D_{IFC} or data containing forensically relevant information D_{IF} or data containing information D_I. Since in this thesis we limit ourselves to data-centric view of the forensic process, the data forming the investigation target D_{IT} are empty. However, the sensor returns data forming the investigation result D_{IR}, thus delivering the data stream for further processing.

The examination step of data gathering for digitised forensics DG_{SF} is in many ways modelled to its counterpart of data gathering for digital forensics DG_{DF} (Section 4.4.1.3). It too serves as the initial data source for the subsequent data investigation and data analysis. And it, too, creates a piece of data that can be seen as a forensic duplicate (see Section 3.3.1.5). This digital representation as a result of the application of the methods of the digitalisation of physical objects DPO and methods for the digitalisation of contextual information DCI (see Section 4.3.2) can be easily digitally copied and subsequent work can be executed on that duplicate. This also means that loss, error and uncertainty (Section 4.1) is only recoverable if the physical source is still available with the desired properties and thus unaffected by short/long-term ageing of substrate and trace and another data gathering run is initiated.

We integrate the part of a three-stage acquisition (data gathering) for substrate data potentially containing trace data as suggested in [HKG+11, p. 5]:

- Profile and survey scans to determine scan parameters and material properties,
- Coarse scans to detect potential regions of interest,
- Detailed scan to gather trace data.

The reason is related both to time and scan/storage resources. Some of the equipment used in our experiments (see also Section 7) requires very precise settings in a very small margin of error in order to gain appropriate results. Thus, in the profile scans, the optimal settings for each region of the substrate are determined. A very low resolution requiring little storage space, the surface of the substrate is surveyed. Obviously, knowledge, experience and data from strategic preparation SP_{SF} based on benchmarking (see Section 4.4.2.1) are invoked. With those results, the surface of the substrate is scanned for potential trace areas (regions of interest) in a higher resolution, thus consuming more storage space and scan time. In a third and most time and storage consuming run, the regions of interest are scanned with a resolution suitable for further processing and subsequent analysis by the human expert (see Section 3.4). All runs can, at least in theory, also employ a pattern recognition pipeline and thus be executed (semi-) autonomously.
All data gathered is stored securely in the chain of custody for digital data.

4.4.2.4 Data Investigation for digitised forensics (DI_{SF})

In the step of data investigation for digitised forensics DI_{SF}, the data gathered from the data gathering step DG_{SF} is investigated for both the trace and from the trace carrier (substrate) as e.g. required by ACE-V (see Section 3.4.1). The following Table 62 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	**Content description based on [KDV15]**	**Context to the digital forensic process**
Data Investigation (DI_{SF})	Includes procedures for the extraction of data from the trace carrier and the trace	Operate on all data types DD as data of the investigation target D_{IT}, return all data types except model data as investigation result D_{IR}, semi autonomous operation for trace specific task such as separation of trace and substrate, separation of overlapping traces, age estimation

Table 62: Description of the examination step of data investigation DI_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during data investigation in digitised forensics DI_{SF}, in reference to the data modelling in Section 4.1 methods of the forensic process SF operate on data containing case-specific, forensically relevant data D_{IFC} or data containing forensically relevant information D_{IF} or data containing information D_I. The aim is also (as in data investigation in digital forensics DI_{DF}) a steady reduction with the goal of having forensically relevant data D_{IFC} or data containing forensically relevant information only.

Data forming the investigation target D_{IT} are start with raw sensor data. Subsequent methods of the forensic process in digitised forensics will work on previously processed data. Only one general exclusion for the data types DD data forming the investigation result D_{IR} can be made, they will never return model data DD_7 (see Section 4.3.2 and Section 4.2.2).

For data investigation in digital forensics DI_{DF} (Section 4.4.1.4), (semi-) autonomous tasks such as resurrection of deleted data, the extraction of meta data etc. are performed, limited to a particular data stream. We derive this pattern for digitised forensics as we include (semi-) autonomous tasks for digitised forensics SF in this step.

From the raw data DD_1 that gathered in the step of data gathering DG_{SF}, contextual data DD_3, parameter data DD_4, (see Section 4.2.2) is extracted (mostly automatically). Further, during the

processing using the pattern recognition pipeline, the intermediate results such as data trace characteristic data DD_5, substrate characteristic data DD_6 are computed and the final result as processed signal data DD_2 is reached. This includes the separation of trace and surface. Some of those tasks, such as the separation of overlapped traces, which only possible in the digital domain see e.g. [QSS+12, pp. 84361A-84361A-9] or trace age detection [Mer14, pp. 1-214] are part of fairly recent and ongoing research.

4.4.2.5 Data Analysis for digitised forensics (DA_{SF})

The examination step of data analysis in digitised forensics DA_{SF}, is concerned with procedures for the separation of trace and trace carrier data (substrate data), its visualisation and subsequent further analysis and annotation. The following Table 63 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	Content description based on [KDV15]	Context to the digital forensic process
Data Analysis (DA_{SF})	Includes procedures for the separation and visualisation of trace data and trace carrier data	Operates on D_{IFC} and D_{IF} with processed signal data DD_2 as data of the investigation target D_{IT}, forensic expert produces annotations containing trace characteristic DD_5 and substrate characteristic feature data DD_6, machine assisted recording of steps taken

Table 63: Description of the examination step of data analysis DA_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during data analysis in digitised forensics DA_{SF}, in reference to the data modelling in Section 4.1 methods of the forensic process SF operate on data containing case-specific, forensically relevant data D_{IFC} or data containing forensically relevant information D_{IF}. The aim is also (as in data analysis in digital forensics DI_{DF}) a steady reduction with the goal of having forensically relevant data D_{IFC} or data containing forensically relevant information only. In this examination step, the examiner operates on processed signal data DD_2 as data of the investigation target D_{IT} and returns trace characteristic feature data DD_5 and substrate characteristic feature data DD_6 as data of the investigation result D_{IR} as part of the annotation process.

The examination step of data analysis DA_{SF} is different compared to its counterpart in digital forensics DA_{SF} in that it is almost exclusively performed by the human forensic examiner. This means that the transfer function underlying the sets of methods for the forensic process in this step are primarily executed by human knowledge and experience as opposed to computing devices. However, the human forensic expert can employ certain facilities on the digital devices operated upon for assistance, e.g. to manually annotate points of interest when comparing traces (typically side by side, see also Section 3.4). An example is the Picture Annotation System PiAnoS [Don20]. So the data processing formalised above is primarily performed by the human expert. However, the recording of steps when using such systems is performed machine assisted and forms the basis for the process accompanying documentations DO_{SF} (see Section 4.4.1.6).

4.4.2.6 Documentation for digitised forensics (DO_{SF})

The examination step of documentation for digitised forensics DO_{SF}, similar to its counterpart in digital forensics DO_{DF} (see Section 4.4.1.6), serves two purposes. During the actual examination every digitally available data is recorded (semi-) automatically (including the strategic preparation) render the examination or strategic preparation comprehensible and some steps repeatable (e.g. investigation steps based on recorded data) by a third person. This is named as *process accompanying documentation*.

After the examination, this data is used to compile the *final report* of the examination. Very likely, multiple variants of such a report will exist that are tailor made for a specific audience, the main difference is likely to be the level of technical detail.

The following Table 64 based on Table 58 contains the description from [KDV15, p. 89] and adds context to the digital forensic process.

Sets of examination steps for digitised forensics SF	**Content description based on [KDV15]**	**Context to the digital forensic process**
Documentation (DO_{SF})	Includes measures for the detailed documentation of the proceedings, also for the "transformation" into a different form of description for the report of the incident	Operates on all data types of digitised forensics SF as data of the investigation target D_{IT} and returns all forensic data types of digitised forensics SF as data of the investigation result D_{IR} for process accompanying documentation, the final report only returns chain of custody data DD_9 and report data DD_{10} as data of the investigation result D_{IR}, forensic data structures in aid for chain of custody for digital data of the examination

Table 64: Description of the examination step of documentation DO_{SF} for digitised forensics SF based on [KDV15, p. 89]

Generally, during the documentation in digitised forensics DO_{SF}, in reference to the data modelling in Section 4.1 methods of the forensic process SF operate on data containing case-specific, forensically relevant data D_{IFC} or data containing forensically relevant information D_{IF}. For process accompanying documentation the data of the investigation target D_{IT}, serving as input sources, comprise all data types of digitised forensics (see Section 4.2.2). The output of methods for the forensic process in digitised forensics SF as data of the examination result D_{IR} again comprises all data types of digitised forensics.

For the final report all data types of digitised forensics can act as data for the investigation target D_{IT}, however, the investigation result will both contain chain of custody data DD_9 and report data DD_{10} as data for the investigation result D_{IR} with the latter likely do contain excerpts of the data collected (often in a pictorial representation).

The step of process accompanying documentation as part of documentation for digitised forensics DO_{SF} is greatly supported by tools recording all data into forensic data storage structures (see appendix Section 10.5) such as the forensic container [KVL11, p. 266] or the FIDEX filing format [For12, p. 1]. This provides a digital chain of custody. Further, the implementation of the concept of provenance (see Section 2.4) is important, which is partly addressed by our Data-Centric Examination approach DCEA. This way the maintenance of the security aspects of integrity and authenticity is supported by the built-in functions and thus such structures can act as a chain of custody for the digital data as results of digitised forensic examination steps. Thanks to the remark of the reviewer Eoghan Casey we point out that in each step a thorough evaluation as outlined in [PCJ+18, p. 7] has to take place, i.e. producing values that can be fed into a decision process. This decision process involves assigning a strength of evidence assignment (see e.g. [Cas19, pp. 1-10]) and needs to be recorded in the process accompanying documentation. Further, the final evaluation, which can also entail event reconstruction, should also manifest itself as a discrete step in itself. However, the latter is out of the scope of this thesis as outlined in Section 1.

Errors, loss and uncertainty is always to be expected, especially if human experts are involved and humans who design the software of the IT systems of the examiners. Thus, in the final report losses, errors and uncertainties must not be covered up but instead explained by the forensic expert instead together with a qualified judgement of their (likely) implications. We believe the Data-Centric Examination Approach DCEA introduced in thus thesis can be great support for those judgements. The main purpose of the final report is the communication of relevant findings

to a variety of audiences (technical, legal personnel, law enforcement, management etc., see [BeC05, p. 151] as described in Section 3.1.9).

4.4.2.7 Flow of the examination process in the context of the examination steps

The examination steps for digitised forensics SF as adapted from digital forensics DF describe actions that are executed by the forensic examiner using sets of methods of the forensic process (Section 4.3.2), which operate on forensic data types for digitised forensics SF (Section 4.2.2). They are connected to one another by providing data for the following step. The following Figure 38 depicts the intended flow of the examination process according to the Data-Centric Examination Process (DCEA).

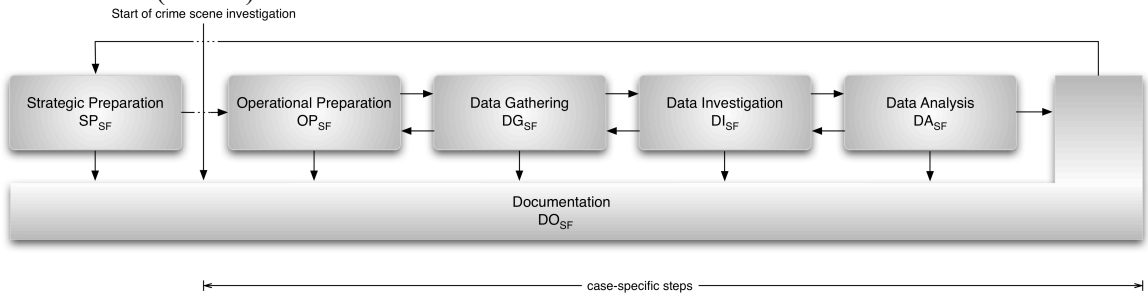

Figure 38: Directional flow of examination steps in digitised forensics with backward steps and the distinct, case-independent step of strategic preparation SP_{SF}, visualised and modified from [KHD09, p. 1] and adapted for digitised forensics SF

Inspired from process described for digital forensics (see Section 4.4.1.7) allow for backward steps during the examination all the way back to operational preparation OP_{SF} since we expect some results in some examinations may lead to a broadening of the search to include previously disregarded or unknown substrates potentially carrying traces for which a dedicated operational preparation is just as necessary. Very important is that each examination step adds to the process accompanying documentation as part of DO_{SF}. Only with a proper and minute recording of all details both the road to the conclusion and the maintenance of the security aspects can a defendable conclusion be achieved. Notice the distinction from strategic preparation SP_{SF} from all remaining examination steps (in Figure 38 using a dotted section within the arrows) since it describes only case-independent actions. The results from the case-specific examination presented in the final report as a result of DO_{SF} and lessons learned are used as feedback to enhance the strategic preparation SP_{SF} in future. The case-specific examinations are triggered by the start of a crime scene investigation.

4.4.3 Reflections and parting thoughts

For both digital forensics DF in Section 4.4.1 and digitised forensics SF in Section 4.4.2 we applied the second part of M3 as part of our methodology outlined in Section 1.2:

> M3:
> We use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics

We looked in digital forensics DF and in digitised forensics SF at means to order the methods for the forensic process from Section 4.3 and the forensic data types the operate upon (Section 4.2) in space and time. We reached the contribution C4 (including C4.1 and C4.2) as presented in Section 1.3:

> C4. Establishment of 6 sets of examination steps for selected use cases:
>
> C4.1 Digital forensics process specific properties of the examination steps based on a systematic review of existing models and selection of best fitting for a data-centric approach
>
> C4.2 Digitised forensics process specific properties of the examination steps based on the application and adaption of the steps from digital forensics

By encapsulating the sets of examination steps into a residual class structure we gave the forensic process a direction as an arrow of time (whilst allowing for backspaces for new findings).

In the following Section 4.5 we reflect on the orchestration of the single contributions to form the Data-Centric Examination Approach DCEA.

4.5 Reflections on loss, error and uncertainty in the context of forensic data types, sets of methods for the forensic process and sets of examination steps and the comparability of forensic examinations

Forensic data types (Section 4.2), sets of methods for the forensic process (Section 4.3) and sets of examination steps (Section 4.4) can form a qualitative means to support the estimation of loss, error and uncertainty (Section 4.1).

Forensic tool testing (e.g. based on [Gray12, p. 1]) as part of strategic preparation (Sections 4.4.1.1 and 4.4.2.1), we propose, is needed to determine knowledge resembling a ground truth. It needs to be known prior to tool use (i.e. a method of sets of methods for the forensic process), which data types form the data of the investigation result D_{IR} when applying a method from the set of methods for the forensic process to given data of the investigation target D_{IR}.

A comparison against the data types for digital forensics DT, and data types for digitised forensics SF alike, actually gathered, investigated, analysed or documented could provide a qualitative estimate for loss, error and uncertainty (as defined in Section 4.1) as proposed in the following:

- loss: any data types not present after processing/transmission (as compared to tool testing) could indicate loss of some sort, i.e. data types from D_{IFC} is not present but can be expected,
- error: data types can be present that are not to be expected after processing/transmission (compared to tool testing), i.e. data types are returned that are outside of D_{IFC} and thus from D_{IF} or D_I,
- uncertainty: data types are the same as expected after processing/transmission, but the concrete data changes with each application of the method, i.e. data types from D_{IFC} are present but with inconsistent values.

Obviously, uncertainty in this context is hardest to detect, as the data types as a qualitative measure do not pinpoint to any anomaly. Re-runs of the methods (if possible) could, however, reveal uncertainty. The evaluation of expected amounts of data (e.g. if tool testing is able to determine a model for expectations) is outside the scope of this thesis and represents future work.

Generally it can be assumed that after ruling out error for each method from the sets of methods for the forensic process involved in a given examination, the forensic examination is likely to be more comprehensive if more data types are contained. By using our methodology from Section 1.2:

M1. We describe loss, error and uncertainty regarding data contained in IT systems based on a modelling of the relationship of all data ever available, data used in all forensic investigations ever and the case-specific data for a given incident

M2. We apply the selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 261] to data stored, processed, communicated in IT systems to distinguish forensic data types for digital and digitised forensics. To achieve this, we construct layers of data by giving them semantics that support the forensic process. In further conjunction with the ISO/OSI model, we use a layering that is not mutual exclusive.

M3. We use residual class based hierarchical approaches (as opposed to a layered non-mutual exclusive description) to define sets of methods for the forensic process, which include tools and toolkits, based on transfer functions derived from [Car03, p. 4]. We further use residual class based hierarchical approaches to define sets of examination steps based on systematic analysis of existing process models from digital forensics. We apply and adapt those examination steps from digital forensics to fit the needs for digitised forensics.

M4. We use forensic data types, sets of methods and sets of examination steps to provide a qualitative estimation on loss, error and uncertainty.

we achieved the following contributions (with added Sections):

C1. Formal description of loss, error and uncertainty using distinct sets of information contained in data (distinction into data containing information D_I, data containing forensically relevant information D_{IFC} and data containing case- specific forensically relevant information D_{IFC} with only the latter solving specific incidents (Section 4.1)

C2. Establishment of a data-centric (data-driven) view in digital and digitised forensics by applying a structured layering of data based on selected characteristics of the ISO/IEC 7498 (commonly known as ISO/OSI reference model) [Cas11, p. 261] in various IT systems (from embedded systems/IoT to data centres) to formulate forensic data types:

- C2.1 Adding semantics to data in the forensic context and establishment of 8 data types for digital forensics for selected use cases (Section 4.2.1)
- C2.2 Adding semantics to data in the forensic context and establishment of 10 data types for digitised forensics for selected use cases (Section 4.2.2)

C3. Establishment of a hierarchical mutual exclusive categorisation of methods for the forensic process:

- C3.1 Distinction into 6 distinct classes based on the likeliness of availability for digital forensics for selected use cases (Section 4.3.1)
- C3.2 Distinction into 10 classes based on the pipeline of the biometric process for a use case (Section 4.3.2)

C4. Establishment of 6 sets of examination steps for selected use cases:

- C4.1 Digital forensics process specific properties of the examination steps based on a systematic review of existing models and selection of best fitting for a data-centric approach (4.4.1)
- C4.2 Digitised forensics process specific properties of the examination steps based on the application and adaption of the steps from digital forensics (Section 4.4.2)

C5. Case-specific qualification of loss, error and uncertainty; and their representation based on forensic data types, methods of the forensic process and investigation steps (Section 4.5)

Generally, we contribute a *common language* (somewhat analogous to [HoL12, p. 1] for computer security incidents) to describe forensic examinations, which provides answers to our research question from Section 1.1:

Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

This common language can render forensic examinations comparable by applying DCEA in an *intra-* and *inter-examination context*. In the intra-examination context, forensic methods can be compared with regards to the data types they operate on (data of the investigation target D_{IT}) and the output (data of the investigation result D_{IR}). Here, the forensic examiner can choose a method based on the provision of data types and the information contained therein deemed necessary for the examination. Further, the whole examination, in theory, can be rendered comparable in an inter-examination context for both the data types collected and examined and regards to the qualitatively estimation of loss, error and uncertainty. DCEA also allows for a graphical representation

of the examination flow as shown with the exemplary use cases in Figure 44, Figure 62 and Figure 67 in the Sections 5.4, 6.4 and 7.4, respectively.

In the following Sections, we will apply the Data-Centric Examination Approach DCEA to selected use cases, all of which are chosen to represent non-standard application fields that we believe to provide challenges for forensic examiners in terms of describing the process with regards to methods used and data operated upon and an estimation of loss, error and uncertainty.

5. Application of the approach to desktop IT Systems used for video surveillance

The contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Tobias Hoppe, Mario Hildebrandt, Claus Vielhauer (in descending order): [KHD09], [KHD+09], [HKD13]

For this use case we pick up the scenario described in [KHD09, p. 1-6], namely that of authentically gathering screen content displayed by an IT system, which is refined in [KHD+09, pp. 1619-1633] and applied to video surveillance and with added gathering of hardware information of the IT system itself off the same data stream. Here, a means is devised to effectively tie displayed media content of an IT system to the respective system and do maintain the IT security aspects of integrity and authenticity starting from the gathering (acquisition) of the data containing the media. The assurance of authenticity is particularly challenging because we operate on digital data only and have no direct physical link to the processing devices as opposed to [HKD13, pp. 86670S-86670S-11], in which we pursue a similar goal (tying physical hardware to data contained in said system) but rely on a physical link between data and device.

We use this use case to show how the application of the Data-Centric Examination Approach DCEA introduced in Section 4 can give us a qualitative estimation of loss, error and uncertainty and a means to describe the forensic process where at the time of writing the original article no universally agreed procedures existed. This is in effect to answer our research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

on the grounds that we can abstractly describe and judge an alternative and sometimes better-suited approach than the then state of the art for tying gathered data to a specific process executed in a specific IT system.

5.1 Special requirements and properties of desktop IT Systems used for video surveillance

As a property it can be assumed that data containing case-specific, forensically relevant data D_{IFC} is to be contained together with forensically relevant data D_{IF} but arguably excluding data containing information D_I outside D_{IF} and D_{IFC} as it is very unlikely that data contained in the programmable IT system that is processed, stored and communicated can be ruled out to be part of those two sets.

We examine an x86 microprocessor based IT system and need software based access to its main memory. For this, strategic preparation SP_{DF} (see Section 4.4.1.1) is essential. The use case requires intentional actions by the operator of the IT system, in our case to place additional software on the file system hard disk drive of the mass storage stream DS_T (see Section 4.2.1) next to the setup of the surveillance system (i.e. the placement of the camera device and support software installation), the email system setup and the triggering of proceedings based on what is visible on the live surveillance picture displayed on the screen of the IT system.

By gathering main memory data on a running system (as opposed to halt and secure memory content off a virtual machine), we will get an inconsistent image (dump) of the main memory stream DS_M (see Section 4.4.1.1) and will alter the resulting data by the very process gathering those data as a structural impact of using such methods (see also Sections 3.3.1.3 and 3.3.1.5).

We focus on essential data (see Section 2.7.2) by both choosing the main memory stream DS_M and on the processes executing the picture creating task ([KHD09, pp. 3-6] and [KHD+09, pp. 1619-1633]) and finally by identifying the components of the IT system used in the picture acquisition ([KHD+09, pp. 1619-1633] only). By choosing the way of executing alternative methods

from the sets of methods for evidence gathering SMG for the step of data gathering DG (see Section 4.4.1.3), we value the order of volatility by first gathering the more volatile main memory as a whole followed by an additional inclusion of swap memory (see Section 4.2.1.2.1) when collecting all memory data of the selected process memory. In doing so, we correlate the main memory data stream DS_M with the mass storage data stream DS_T, where the swapfile resides as a regular file in the filesystem. Only by exploiting the fact that the operating system at the time knows, which portion of the swapfile belongs to which process, this data can be tied to the process structure in main memory, giving it more evidentiary value (see Section 2.7.2).

We use methods of the operating system OS (see Section 4.3.1) to enumerate hardware and store its results in main memory and in this thesis suggest even more essential data for hardware identification as compared to [KHD+09, pp. 1619-1633]. Here we exploit the fact that the operating system to enumerates the hardware as essential data (see Section 2.7.2) and stores in on the Microsoft Windows registry in main memory.

For picture recovery we use methods for the forensic process (see Section 4.3.1) of the operating system OS and methods of the IT application ITA based on the fact that for images to be displayed on the Screen of the IT system, it has to be displayed as a bitmap, which is stored in process memory and can be extracted thereof.

We employ an external time stamping service for both integrity and authenticity assurance for the logs and thus the process accompanying documentation as part of the examination step of documentation DO (see Section 4.4.1.6).

5.2 Exemplary chosen use case - Examination of main memory

In our case as described in [KHD+09, p. 1619], a webcam video camera is attached to a laptop PC and the webcam is overlooking a parking lot. Especially maintaining authenticity is a challenging using conventional means (e.g. photographing the screen content with a digital camera as suggested by e.g. [ACPO20, p. 12] or by only looking at the data stored on the mass storage stream, i.e. the filesystem on the harddisk drive). Our goal in [KHD09, p. 3] and its revision in [KHD+09, p. 1619] is to show the extraction of screen content that was arguably on an IT system's display at some point in time because its video data is contained in a running process. We want to ensure (wherever possible, sometimes using makeshift but justified approaches) integrity and authenticity of all data gathered, investigated, analysed and documented.

In this thesis we add another approach to tie data to its processing hardware, which we believe gathers more essential data and thus of higher evidentiary value. The IT system used in the original articles [KHD09, pp. 1-6] and [KHD+09, pp. 1619-1633] is decommissioned by now but the data acquired is still available. We will use this remaining data and the original description of proceedings to show our new idea. Generally, the level of detail (see Section 4.4.1.2) is set as such that we treat a desktop IT system as an atomic unit and not looking into the internal structure within those devices.

5.3 Practical tests

We now use our method for ordering forensic data types (see Section 4.2.1) and sets methods for the forensic process (see Section 4.3.1), i.e. the sets of examination steps (see Section 4.4.1), to describe the flow of proceedings in the forensic examination of our exemplary chosen scenario analogous to the generic description in Section 4.4.1.7. We also use the data streams DS defined in the description of forensic data types (see Section 4.2.1). We offer a sort of high-level view on our experiments tuned to the requirements of our Data-Centric Examination Approach DCEA introduced in Section 4 of this thesis. For a more detailed description, we direct the interested reader to the original articles [KHD09, pp. 3-6] and [KHD+09, pp. 1621-1633].

We perform an exemplary evaluation of loss and uncertainty on the forensic data type of hardware data DT_2 for digital forensics DF and apply the set relationships regarding this forensic data type using the Venn diagram representation from Sections 4.

5.3.1 Strategic preparation SP_{DF} for the use case of the examination of main memory

For our description of the strategic preparation we select items from Section 4.4.1.1 that we deem relevant for our purpose of the exemplary examination as described in Section 5.2. The whole use case is only possible due to strategic preparation; we need the intention and cooperation of the operator of the system to be able to conduct the examination in the first place. We start by performing a system landscape analysis as depicted in the following Figure 39, which we provide additionally to the findings in [KHD+09, pp. 11619-1633].

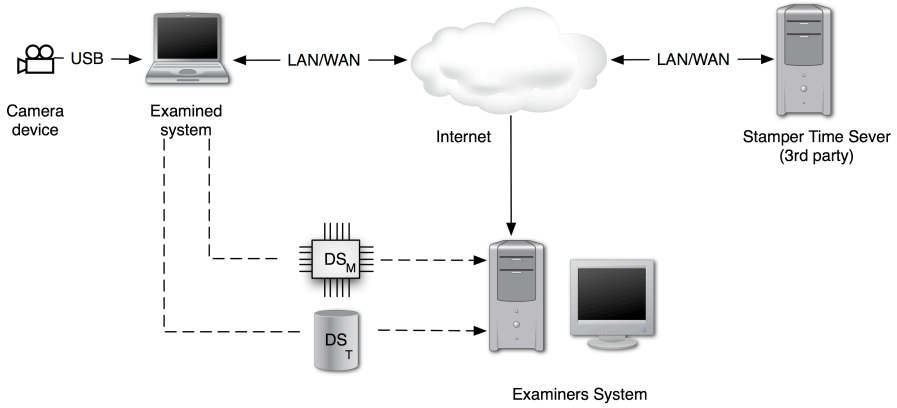

Figure 39: System landscape analysis as part of strategic preparation for digital forensics SP_{DF} for the use case of camera surveillance with system under examination and the examining system and a third party service provider and the data stream of mass storage DS_T and the data stream of main memory DS_M depicted as a dashed line

The examined system is connected to the network to support the maintenance of the security aspects of integrity and authenticity only and thus the network stream DS_N itself is not a target of the examination but a tool. However, in the true spirit of a forensic examination, where the proceedings itself are not hidden way but declared, we have to enumerate that data stream also. The primary focus is on the main memory data stream DS_M but we use a very small part of the mass storage stream DS_T as well as we are accessing the virtual memory where portions of it may be resident on the swap file stored on mass storage (see Section 4.2.1.2.1). The lines regarding DS_M and DS_T are dashed in Figure 39 because they do not represent a permanent connection but the origins of data to be investigated, analysed and documented on the examiners IT system. We also prepare an inventory of the software installed on the examined system, which is displayed in the following Table 65:

Software on the examined system	**Exploited sets of methods for the forensic process**
Microsoft Windows XP Service Pack 3 including msinfo32	OS ITA
SecureCam	ITA
win32dd	OS
sha256sum	DPE
blat	ITA
ProcMemDump	OS
SecureDump	OS

Table 65: Software installed on the examined system as part of strategic preparation in digital forensics SP_{DF}

We look at the methods we use from those installed software items based on the sets of methods for the forensic process in digital forensics DF (Section 4.3.1). The system has Windows XP installed as the operating system and we will use it as a provider for methods for the forensic process from the operating system OS. We install SecureCam [Bed20] as our IT application to pro-

vide the camera pictures. We use the command line E-Mail tool blat [Mus20] as part of the methods of the IT application ITA to connect to the Stamper timeserver [Ric20]. We install win32dd [Sui20] and our own ProcMemDump as a means for data acquisition. Both access methods of the operating system OS for the acquisition. We further install the cryptographic hashsumming tool sha265sum [Pri20] from the sets of data processing and evaluation DPE methods to maintain the integrity of the data processed and stored and the hardware information display application msinfo32 as an IT application ITA. Our own wrapper application SecurceDump, which calls the aforementioned software in an orchestrated manner, since it is only invoked if the examination is triggered, can be categorised as a method for the scaling of methods for evidence gathering SMG from the sets of methods for the forensic process (Section 4.3.1).
Further, we install the software for the examiners system as follows (Table 66):

Software on the examiners system	**Exploited sets of methods for the forensic process**
Microsoft Windows XP SP 3	OS
IrfanView	ITA
Volatility	DPE
sha256sum	DPE
Microsoft Excel 2003	ITA
Python	ITA
HxD	DPE
Bitmap finder	DPE

Table 66: Software installed on the examiners system as part of strategic preparation in digital forensics SP_{DF}

The system has Windows XP installed as the operating system, which provides methods of the operating system OS. A forensic suite in broad active use for memory data stream DS_M examination is the Volatility toolset [Vol20], which we thus use and classify as a provider of methods for data processing and evaluation DPE. Also acting as DPE for the verification of the integrity we install the sha256sum application [Pri20] and for an unobstructed view on all symbols contained in a file we install the hexadecimal editor HxD [Hör20] and for the piecewise entropy calculation a self coded program called Bitmap finder. As an integral part of our picture analysis we will use the IrfanView [Ski20] image processing suite and the python interpreter [Pyt20] (required for Volatility) and the Microsoft Excel 2003 application, all of which we classify as IT application ITA with respect to the methods used.

All of those case-independent actions are documented as part of the process accompanying documentation as part of the documentation DO_{DF} step (see Section 5.3.6). With those preparations we are sorted for the invocation of a forensic examination.

5.3.2 Operational preparation OP_{DF} for the use case of the examination of main memory

During operational preparation OP_{DF} we provide for sanitized storage media (see Section 3.3.1.6) for the data transfer from the examined IT system to the examiners IT system and initiate a chain of custody for digital objects by the provision of a structure protected storage space on the mass storage of the examiners system. We deem the examiners IT system fit for purpose based on the software installed according to Table 66 from Section 5.3.1. All of those case-specific actions are documented as part of the process accompanying documentation DO_{DF} step (see Section 5.3.6). The level of detail (see Section 4.4.1.2) is set as such that we treat a desktop IT system as an atomic unit and not looking into the internal structure within those devices.

5.3.3 Data gathering DG_{DF} for the use case of the examination of main memory

We summarise the step of data gathering DG_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited

set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 67:

Specific instantiation of the method of the forensic process	Accessed data streams DS	Forensic data type for data of the investigation target D_{IT}	Exploited method of the forensic process	Forensic data type for data of the investigation result D_{IR}	Involved sets regarding the information contained
win32dd	DS_M	DT_1	OS	DT_1, DT_3	D_{IF}, D_{IFC}
ProcMemDump	DS_M, DS_T	DT_1	OS	DT_1, DT_3, DT_6	D_{IF}, D_{IFC}
MSInfo	DS_M	DT_1	ITA	DT_1, DT_2	D_{IF}, D_{IFC}

Table 67: Summary of the actions taken during data gathering DG_{DF} based on the findings of [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

During data gathering DG_{DF} we try to access the data in its lowest and most encapsulated form that is raw data (see also Section 4.2.1) and thus try to access our data from the investigation target D_{IT} at the lowest and most comprehensive way by exploiting methods of the forensic process offered by the operating system OS. In [KHD09, pp. 3-4] and refined in [KHD+09, pp. 1625-1627] we devise two different strategies to access the main memory of the examined IT system. In the order of volatility (see Section 3.3.1.1) we first use the method offered by win32dd and exclusively reading the main memory data stream DS_M to create a dump of all main memory of the whole examined IT system; it thus equals the amount of physical RAM available and includes kernel and user space data. The method simply writes all captured data in a file on the file system of the examined system as the data of the investigation result D_{IR}. Since no further processing takes place, this data still represents raw data DT_1. Additionally, we get an output of meta data about the data gathered as details about data DT_3. We disregard them for our experiments but document their presence during process accompanying documentation.

Thereafter use the method offered by our own implementation ProcMemDump involving the functions VirtualQueryEx and Read Process Memory to create a dump of virtual memory of the process running the SecureCam application. Since we are gathering virtual memory, next to the main memory data stream DS_M we also collect swapped out memory residing in the mass storage data stream DS_T (see Section 4.2.1.2.2). The method simply writes all captured data in a file on the file system of the examined system as the data of the investigation result D_{IR}. Since no further processing takes place, this data still represents raw data DT_1. However, we are also presented with a list of existing processes to select the process of interest together with extra data (e.g. Process ID, starting Process ID, memory allocated etc.) corresponding to details about data DT_3 and process data DT_6 for the main memory data stream DS, respectively. We disregard them for our experiments but document their presence during process accompanying documentation.

In [KHD+09, pp. 1631-1632] we use the method provided msinfo32 as part of the sets of methods provided by IT application to query the hardware of the examined system. Calling this tool before the data gathering ensures that hardware information is contained in the kernel space dump of the main memory data stream DS_M. Msinfo32 as a method from the sets of methods provided by IT applications ITA operates on raw data DT_1 as data of the investigation target D_{IT} and provides hardware data DT_2 in a normal run of the application as the data of the investigation result D_{IR}. However, in an effort to encapsulate the data in the singular kernel space main memory dump, we disregard the processed data and rely on result data being present in the raw data DT_1 for further analysis in the data analysis step DA in the next Section 5.3.5. We disregard DT_2 for our experiments but document their presence during process accompanying documentation.

Since we are using software-based methods to capture the dump that is running on the respective system and the creation of the dump will take some time while the system is still running, our data will be inconsistent with the data present at the time t_i at the start of the gathering process. This inconsistency is inherent in the process but we must not hide it but rather explain its origins. Naturally, we try and minimise the effect as best as possible by not starting any new processes and shift large amounts of data within the main memory. We expect to have captured as many as possible volatile data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}.

All of those case-specific actions are documented as part of the process accompanying documentation as part of documentation DO_{DF} (see Section 5.3.6).

5.3.4 Data investigation DI_{DF} for the use case of the examination of main memory

We summarise the step of data investigation DI_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 68:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
volatility	DS_M, DS_T	DT_1	DPE	DT_1, DT_3, DT_6	D_{IF}, D_{IFC}
IrfanView	DS_M, DS_T	DT_1	ITA	DT_8	D_{IF}, D_{IFC}

Table 68: Summary of the actions taken during data investigation DI_{DF} based on [KHD+09, pp. 1627-1632] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

In the examination step of data investigation DI_{DF} we use the data gathered in the previous step and concentrate on a reduction of the amount of data and reducing the amount of data that is exclusively in the set of data containing forensically relevant information D_{IF} whilst maintaining the amount of data from the set of data containing case-specific, forensically relevant information D_{IFC}. Most of the steps operate (semi-) automatic and do not require substantial prior case-specific knowledge (see also Section 4.4.1.4). Note that we define the data stream DS the methods are operating on according to its origin and not according to the intermediate representation (i.e. the file containing the dump on the mass storage device of the examiners IT system). Thus, whenever we use data from the user space dump (gathered using the ProcMemDump method) accessing virtual memory, chances are that we are also including data from the mass storage data stream DS_T due to the swapping mechanism (see Section 4.2.1.2.2).

For the picture extraction process on the user space dump from using the method ProcMemDump in data gathering DG_{DF} in data investigation DI_{DF} we use a method provided by volatility from the sets of methods for data processing and evaluation DPE to add structure to the raw data, namely by sorting the data into data areas belonging to specific processes (see also [LCL+14, pp. 149-160]). It uses the raw data DT_1 containing the main memory data stream DS_M as data of the investigation target D_{IT}. With volatility we retrieve the process list as part of process data DT_6 and meta information about the memory dump as details about data DT_3. Mostly we are interested in the data area that belongs to the process initiated by the SecureCam application. That data area is saved as a regular file in the filesystem of the examiners IT system. In our case that area constitutes raw data DT_1 since we are not giving it semantics, yet. Those DT_1, DT_3 and DT_6 form our

data of the investigation result D_{IR}. We will proceed with DT_1 in the data analysis step DA in the next Section 5.3.5. We disregard DT_6 for our experiments but document their presence during process accompanying documentation.

For the picture extraction process on the raw kernel space dump gathered with the method win32dd during the data gathering step DG_{DF} in the data investigation step we use the method provided by IrfanView application from the sets of methods of the IT application ITA. It uses the raw data DT_1 containing the main memory data stream DS_M as data of the investigation target D_{IT}. Here we have access to kernel-space picture data in de-allocated, i.e. memory pages marked as deleted, albeit in a limited fashion (see also Section 4.2.1.2.2). As stated in [KHD+09, pp. 1631], we are limited to 4096 consecutive bytes as a typical size for memory pages. We can use IrfanView to display an arbitrary number of bytes as a bitmap picture, if we use fitting settings (we direct the interested reader to [KHD+09, pp. 1619-1633] for details). Thus our data of the investigation result D_{IR} are pictures and as such they represent user data DT_8.

Diverting from the proceedings described in [KHD+09, pp. 1619-1633] on the basis of a refined process, in this thesis we suggest to use hardware data DT_2 maintained in the windows registry. To access it, we use a method provided by volatility from the sets of methods for data processing and evaluation DPE. It uses the raw data DT_1 containing the main memory data stream DS_M as data of the investigation target D_{IT}. With volatility we retrieve the windows registry keys present in main memory that contain hardware data DT_2 as data of the investigation result D_{IR}, namely vendor IDs VID, product IDs PID and serial numbers. We choose this new approach since this data is believed to be of a higher evidential value than the previous approach of arbitrary looking for hardware describing strings somewhere in the kernel space memory dump. We reason that a modification of hardware data DT_2 in the registry could lead to system malfunctions as it is queried by drivers and other software. Thus we treat this data as essential data (see Section 2.7.2).

Since we did not publish those proceedings before, we go into a little more detail here whilst maintaining brevity. We identify the region of the registry using the hivelist command of the volatility method and get pointers to virtual and physical memory addresses together with the name associated with that part of the registry structure (Figure 40).

```
Virtual    Physical   Name
---------- ---------- ----
0xe22854a0 0x130a14a0 \Device\HarddiskVolume1\Dokumente und Einstellungen\a1\Lokale Einstellungen\Anwendungsdaten\Microsoft\Windows\UsrClass.dat
0xe1df6758 0x12857758 \Device\HarddiskVolume1\Dokumente und Einstellungen\a1\NTUSER.DAT
0xe1eae718 0x129c0718 \Device\HarddiskVolume1\Dokumente und Einstellungen\LocalService\Lokale Einstellungen\Anwendungsdaten\Microsoft\Windows\UsrClass.dat
0xe1ea1b60 0x128bcb60 \Device\HarddiskVolume1\Dokumente und Einstellungen\LocalService\NTUSER.DAT
0xe1b9bb60 0x1da83b60 \Device\HarddiskVolume1\Dokumente und Einstellungen\NetworkService\Lokale Einstellungen\Anwendungsdaten\Microsoft\Windows\UsrClass.dat
0xe1bb7b60 0x1223eb60 \Device\HarddiskVolume1\Dokumente und Einstellungen\NetworkService\NTUSER.DAT
0xe1693b60 0x0f4c0b60 \Device\HarddiskVolume1\WINDOWS\system32\config\software
0xe1a7cb60 0x0eb7bb60 \Device\HarddiskVolume1\WINDOWS\system32\config\default
0xe169f008 0x0f4d0008 \Device\HarddiskVolume1\WINDOWS\system32\config\SAM
0xe1693008 0x0f4c0008 \Device\HarddiskVolume1\WINDOWS\system32\config\SECURITY
0xe1576008 0x02d83008 [no name]
0xe1035b60 0x0282fb60 \Device\HarddiskVolume1\WINDOWS\system32\config\system
0xe1022758 0x02806758 [no name]
```

Figure 40: Screenshot of the hivelist method from the volatility tool suite [Vol20] containing the virtual address of the system configuration entries of the Windows registry (in our experiment 0xe1035b60)

We then pick the appropriate part of the registry (in our experiment the system configuration at virtual address 0xe1035b60) and dump its content using the hivedump command of the volatility method into a regular file on the filesystem of the examiners IT system as details about data DT_3 for further analysis in the data analysis step DA_{DF} in the next Section 5.3.5.

All of those case-specific actions are documented as part of the process accompanying documentation as part of documentation DO_{DF} (see Section 5.3.6).

5.3.5 Data analysis DA_{DF} for the use case of the examination of main memory

We summarise the step of data analysis DA_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types

DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 69:

Specific instantiation of the method of the forensic process	Accessed data streams DS	Forensic data type for data of the investigation target D_{IT}	Exploited method of the forensic process	Forensic data type for data of the investigation result D_{IR}	Involved sets regarding the information contained
bitmap finder	DS_M, DS_T	DT_1	DPE	DT_3	D_{IF}, D_{IFC}
Microsoft Excel 2003	DS_M, DS_T	DT_3	ITA	DT_3	D_{IF}, D_{IFC}
HxD	DS_M	DT_1, DT_3	DPE	DT_2	D_{IFC}
Human examiner	DS_M, DS_T	DT_8	DPE	DT_8	D_{IFC}

Table 69: Summary of the actions taken during data analysis DA_{DF} based on [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

In the examination step of data analysis DA_{DF} we use the data investigated in the previous step and concentrate on a further reduction of the amount of data and reducing the amount of data that is exclusively in the set of data containing forensically relevant information D_{IF} whilst maintaining the amount of data from the set of data containing case-specific, forensically relevant information D_{IFC}. This step relies heavily on prior case-specific knowledge from the previous steps and is only assisted by the methods described in the following. We highlight this by adding the methods for the forensic process provided by the examiner as sets of methods for data processing and evaluation DPE (see Table 69). This is due to the experimental nature of the examination of the use case as research work and is in many ways comparable to the data analysis step in digitised forensics DA_{SF} (see Section 4.4.2.5). We rely on backward steps to the data investigation DI_{SF}.

For [KHD09, pp. 1-6] and [KHD+09, pp. 1619-1631] the examiner identifies and analyses the result of the content loaded into IrfanView as data of the investigation target D_{IT} from data investigation DI_{DF} (Section 5.3.4) as the forensic data type of user data DT_8 and identifies and secures case-specific, forensically data D_{IFC} as the pictures as user data DT_8 as data of the investigation result D_{IR}. Only the examiner in this step can add semantics to the picture content under analysis.

Assisting the step is the self-written method of bitmap finder as method of data processing and evaluation DPE, which computes the entropy over a sliding window operating on the raw data DT_1 as the data of the investigation target D_{IT}. Data of the investigation result D_{IR} are meta data of the entropy over a given segment and constitute details about data DT_3. We use the spreadsheet application Microsoft Excel 2003 as method of data processing and evaluation DPE from the sets of methods of the forensic process for digital forensics DF to process and display the information contained in those details about data DT_3 representing the data of the investigation target D_{IT}. Data of investigation result D_{IR} are meta data acting as pointers towards picture content in the dump and constituting details about data DT_3. Each finding triggers a back step towards data investigation DI_{DF} using IrfanView. The assistance functionality reduces but does not completely removes non-case specific data containing forensically relevant information D_{IF}.

In [KHD09, pp. 1-6] we do not extract any hardware data DT_2 in order to tie the data to a particular piece of hardware. The absence of DT_2 can serve as a first qualitative indicator for *loss* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA, (see also Figure 41 for the Venn diagram representation).

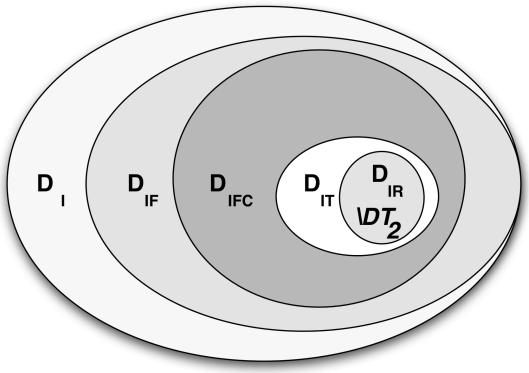

Figure 41: Loss due to the absence of hardware data DT_2 (denoted by the "\" sign) as data of the investigation result D_{IR}

In [KHD+09, p. 1622] we use the hexadecimal editor HxD read-only as a method from the sets of methods for data processing and evaluation DPE to search for strings contained the dumps of full physical memory dump as a result of data gathering DG_{DF} (Section 5.3.3) and for strings contained in the extracted process memory from data investigation DI_{DF} (Section 5.3.4). Both are used as raw data DT_1 acting as data of the investigation target D_{IT}. The string search confirms the presence of hardware data DT_2 in the respective dumps as data of the investigation result D_{IT}, which represent data containing case-specific, forensically relevant data D_{IFC} only. This data is placed in there by the usage of the msinfo32 method in data investigation DI_{DF} (Section 5.3.4).

In [KHD+09, p. 1622] we extract hardware data DT_2 from non-essential data. This is easily forged by placing other hardware information in main memory, which can serve as a first qualitative indicator for *uncertainty* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA, (see also Figure 42 for the Venn diagram representation).

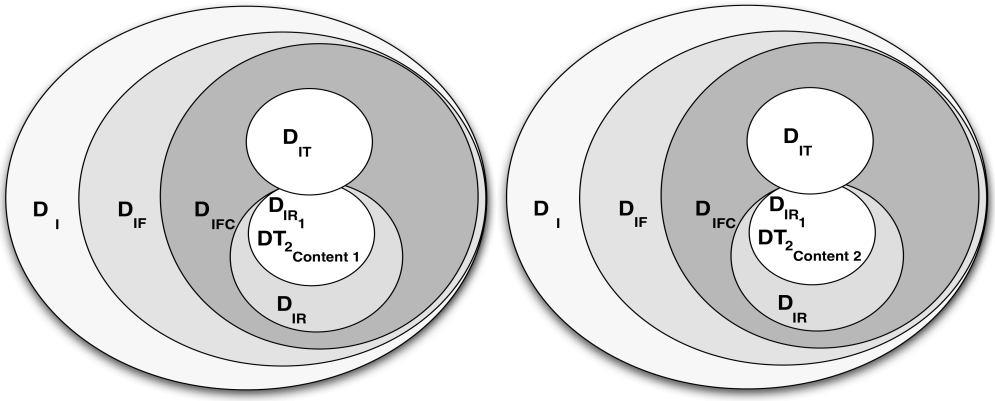

Figure 42: Uncertainty (depicted as content 1 left and content 2 right) due to hardware data DT_2 not being extracted from essential data and thus falling victim to forgery easily

In reviewing the proceedings published in [KHD+09, pp. 1619-1633] we realise that the hardware data DT_2 collected in this fashion are not strictly essential data and thus of a lower evidentiary value (see Section 2.7.2). In this thesis we look for different means and use the strategy involving the analysis of the hardware information section from the registry (see Section 5.3.4). We use the identified hivedump and load them as details about data DT_3 acting as data of the investigation target D_{IT} into the HxD hexadecimal editor in read-only proceedings as a method of data processing and evaluation DPE. We look for vendor IDs (VID), product IDs (PID) and serial numbers of the devices contained, acting as hardware data DT_2 as data of the investigation result D_{IT} (Figure 43).

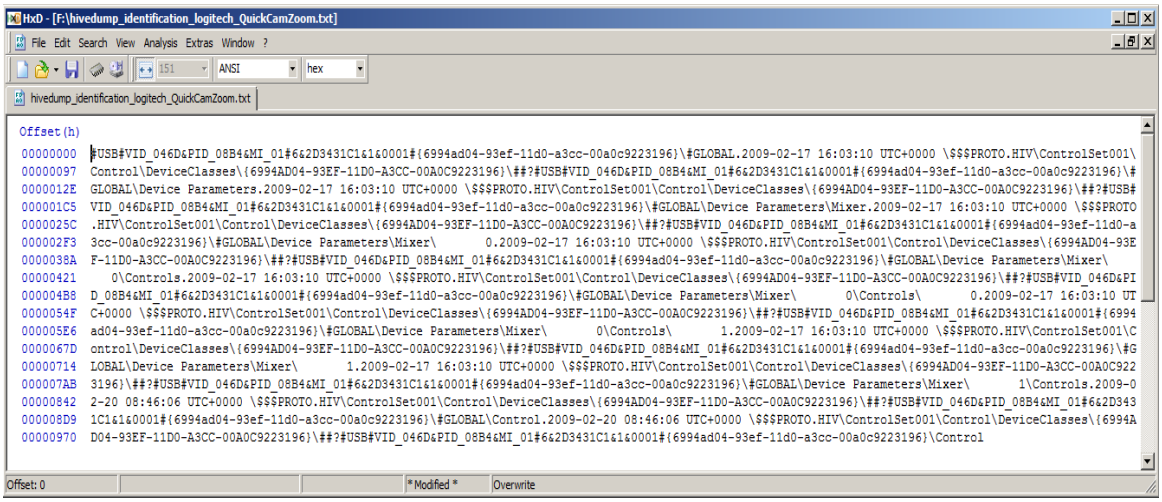

Figure 43: Excerpt of the hivedump viewed with HxD [Hör20] displaying the device with the VID 046D, PID 08B4 and serial number 2D3431C1 that represents a USB attached Logitech QuickCam Zoom video camera according to [Thw20]

Existing online data bases such as [Thw20] identifying components based on the VID and PID can be of great help in identifying relevant hardware components, i.e. those that form the core and are not easily replaced (e.g. for laptops mainboard, graphics card etc.). This way we expect to gather more essential data as the registry is essential for the functionality of a Microsoft Windows operating system. We can reconstruct the hardware inventory of the examined system in a fashion that we argue is of higher evidentiary value.

All of those case-specific actions are documented as part of the process accompanying documentation as part of documentation DO_{DF} (see Section 5.3.6).

5.3.6 Documentation DO_{DF} for the use case of the examination of main memory

We summarise the step of documentation DO_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 70:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
sha256sum	DS_M, DS_T	DT_1, DT_2, DT_3, DT_6, DT_7, DT_8	DPE	DT_3	D_{IF}, D_{IFC}
blat	DS_M	DT_7	ITA	DT_3	D_{IF}, D_{IFC}

Table 70: Summary of the actions taken during the examination step of documentation based on [KHD+09, pp. 1622-1627] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

Across the whole examination, each and every piece of data needs to be recorded to maintain a digital chain of custody for digital forensics (see Section 2.4) process accompanying documentation as part of the examination step of documentation DO_{DF} (Section 4.4.1.6). This data, by its

very nature involves the set of data containing forensically relevant information D_{IF} and the set of data containing case-specific, forensically relevant information D_{IFC}. To maintain and verify its integrity, the method sha256sum from the methods of data processing and evaluation DPE is used to compute a cryptographic hash sum. Note, that hash calculation algorithms are subject to be compromised by malicious collisions (i.e. different input data generating the same hash sum), thus the most recent and, at the time, known to be collision free algorithm should be chosen. All data types of the whole examination in their file representation act as data of the investigation target D_{IT}. The method computes the cryptographic hash sum acting as data of the investigation result D_{IR} of the data type of details about data DT_3. In [KHA+09, p. 1627] we suggest using an external trusted, signing time service called Stamper [Ric20]. This is, admittedly, a makeshift solution but we include it for the sake of completeness. The log files of all automated activity are sent to that signing and time stamping service using the command line email program blat [Mus20]. It supplies methods from the set of methods of IT applications ITA and uses session data DT_7 of the forensic examination as data of the investigation target D_{IT} and returns the stamp (i.e. meta) data as data of the investigation result D_{IR} as details about data DT_3.

For all intents and purposes, this description of proceedings of the whole Section 5 constitutes a final report as part of the examination step of documentation DO_{DF} as described in Section 4.4.1.6.

5.4 Summary of new findings and evaluation with regards to loss, error and uncertainty for desktop IT systems used for video surveillance in the use case of examination of main memory

In summary we can constitute that the examination of the use case can be described using our forensic data types DT for digital forensics DF from Section 4.2.1, the sets of methods for the forensic process in digital forensics DF from Section 4.3.1 and the sets of examination steps for digital forensics DF from Section 4.4.1. We can use the separation of data of the investigation target D_{IT} and the data of the investigation result D_{IR} and the set relations between data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}. The Figure 44 summarises the examination.

This description provides *common language* to comparably describe different forensic examinations in a structured way, thereby allowing to systematically address issues of loss, error and uncertainty within procedures.

When describing the use case of a memory examination in IT systems used for video surveillance using the Data-Centric Examination Approach DCEA, we can also use the data types for a first indicator for loss, error and uncertainty in a qualitative estimation as suggested in Section 4.5. We will separate the three approaches used within the description of the process starting from Section 5.1 onwards starting with our approach from [KHD09, pp. 1-6], followed by our approach from [KHD+09, pp. 1619-1633] and we close with our enhancements presented in this thesis.

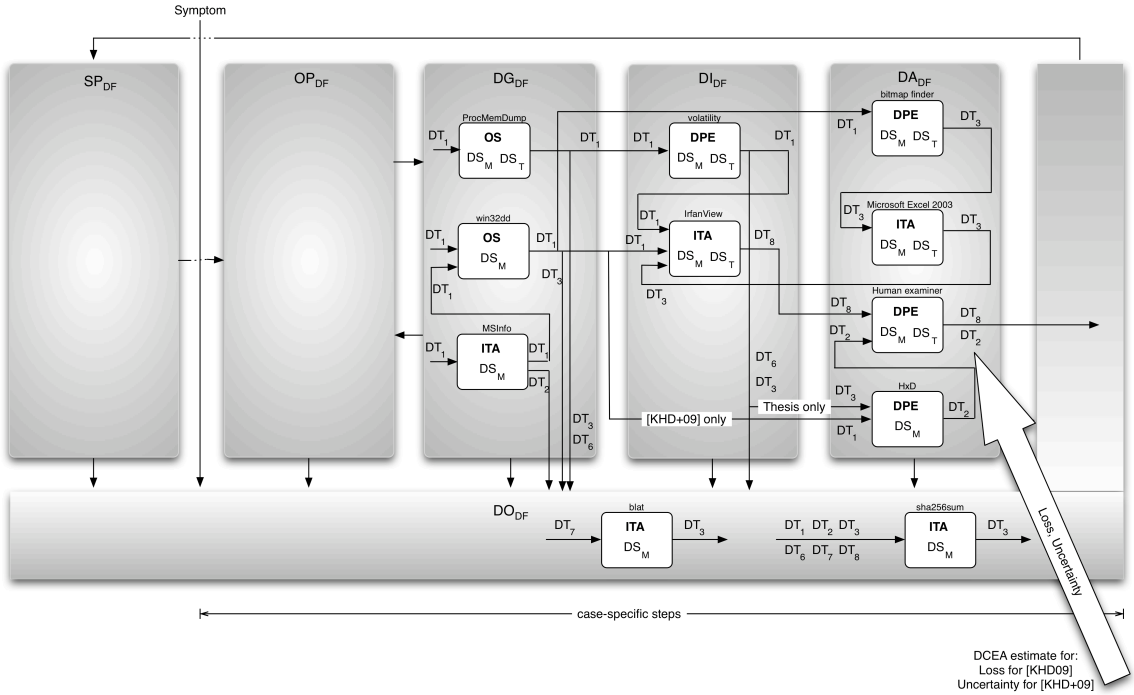

Figure 44: Depiction of the examination flow as summary of the use case of examination of main memory with the involved forensic data types and sets of methods for the forensic process

In the picture recovery we cannot constitute *loss, error* or *uncertainty* according to the qualitative measure from our Data-Centric Examination Approach DCEA using the forensic data types DT for both data of the investigation target D_{IT} and data of the investigation result D_{IR}. We received all expected D_{IR} and worked on all expected D_{IT}.
However, regarding our stated goal of having means to tie the picture data to a specific piece of video surveillance hardware (similar goal to [HKD13, pp. 86670S - 86670S-11] but entirely different means), for the approach from [KHD09, pp. 1-6] we can constitute an obvious *loss* since we do not gather, investigate or analyse the forensic data type of hardware data DT_2.

For the approach of [KHD+09, pp. 1619-1633] we can constitute *uncertainty*, since the places used to recover hardware data are non-essential (see Section 2.7.2) and could contain other information based on other processing performed on the IT system or on information entered by a malicious operator.

In our third approach we can argue to rely on essential information and thus reducing this uncertainty with regards to the hardware data DT_2 retrieved from the examined system. In Section 6 we show, how hardware data can still be the cause for uncertainty and we need to enhance the level of detail (see Section 4.4.1.2) to start tackling this challenge.

In all three approaches we can speculate about both loss and uncertainty since we are using software-based methods to capture the dump that is running on the respective system and the creation of the dump will take some time while the system is still running, our data will be inconsistent with the data present at the time t_i at the start of the gathering process (see also Section 5.3.3). However, in our tests this inconsistency did not manifest itself in the qualitative measure provided by our Data-Centric Examination Approach DCEA. Here a quantitative measure left for future work could be forthcoming.

Further we can see future work in connecting digital forensics DF and digitised forensics SF by integrating semantic contents of the camera images themselves, involving digitised forensics and a description of proceedings, accordingly (see also Section 9.4).

6. Application of the approach to processor-controlled components

Some of the contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Mario, Hildebrandt, Robert Altschaffel, Tobias Hoppe, Kevin Lamshöft (in descending order): [HKD08], [ALK+18], [AHK+19]

Since this use case contains our research that is not published otherwise, we describe it in more detail compared to the use cases based on our published articles. For this use case we pick up a problem that is discussed in the forensics community for some time but still no universally agreed, fully documented solution is known to us. We focus on hidden storage and device impersonation [Pol20, p. 3] functionality enabled components that make up IT systems (e.g. desktop IT, automotive devices, industrial control systems, home automation, Internet of things etc.) and which we name as processor-controlled components for the remainder of this thesis. This problem is highlighted e.g. in [Dom19], [KaK20, pp. 1-81], [RXS+13, pp. 281-286], [NYE17, pp. 675-688], [Kas19, pp. 1-44], [Ecl20], [Del20, pp. 1-140], just to name a few.

In this thesis for our practical tests we will limit ourselves to external mass storage devices. The underlying problem goes further than the computer forensics tool testing programme CFTT regarding disk imaging subtask [NIST20], which is concerned with verifying whether the number of sectors reported fits the data gathered by the tool. In the following we are going to question, whether the storage layout and content reported by mass storage devices as one instance of processer-controlled components is actually true or if there is potential for data hiding and hidden functionality. Further we look at means to perform device impersonation [Pol20, pp. 38-55], i.e. whether we can change hardware data such as the serial number, the vendor ID (VID) or the product ID (PID) or the device names at different interface abstractions to arbitrary values.

This can be used for malicious purposes and anti-forensics (see Section 2.6); an attacker can borrow a (removable) device from a victim, record the hardware data and forge a device with those data and thus divert (at least initial) attention to that original device owner (e.g. in case of a malware attack staged afterwards). Or the owner of a USB thumb drive that is used to stage an attack using malware maintains the content (the malicious software) but changes the hardware information data afterwards in the expectation of the seizure of the device in the proceedings of the examination of the IT system and subsequent detection of removable storage media. In a third and completely different context, outlined in [Pol20, p. 3], this information could be used as part of bypassing endpoint security based on this data containing hardware information, allowing for unauthorised data injection or extraction that is very difficult to prove at best, as we will show.

We will apply our Data-Centric Examination Approach from Section 4 at a higher level of detail (see Section 4.4.1.2) with the goal of qualitatively estimating loss, error and uncertainty where at the time of writing of this thesis to our knowledge no universally agreed procedures existed. This is in effect to answer our research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

In addressing this question, first we have to take a closer look on the innards of a conventional IT system as used in desktops and servers. Much of what is said also applies to non-standard devices e.g. for usage in automotive environments (see also [ALK+18, pp. 104-117]) or industrial control systems ICS (see also [AHK+19, pp. 128-136].

6.1 Special requirements and properties of processor-controlled components

We can safely assume the absence of data that is only data containing information D_I but not data containing forensically relevant information D_{IF} or data containing case-specific, forensically relevant information D_{IFC} (see Section 4.1).

To get a first idea of the importance of the examination of processor-controlled components we take a detailed look at a common desktop IT system (see Figure 45). It is by no means an atomic unit but a networked complex IT system in itself. We deliberately choose a system that is in use for quite some time on the grounds that generally the forensic expert needs to be prepared for any IT environment and that those systems are still in use nowadays but will reflect on changes posed by new developments.

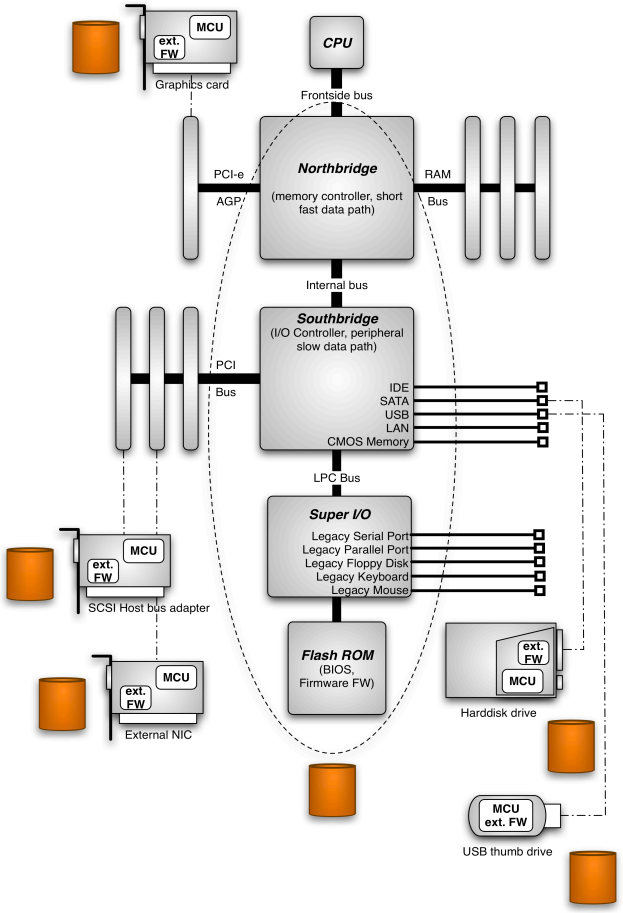

Figure 45: Simplified desktop IT system as a network of processor controlled components

As can be seen, apart from the visible CPU, there are numerous other components that are processor-controlled with own RAM, ROM and bus access. Thus, according to the higher level of detail selection, as discussed in Section 4.4.1.2, we see a multitude of forensic data sources (depicted as orange barrels), each potentially with their own mass storage, main memory and networked stream (if applicable) and their own processing unit, often realised as a System-on-Chip SoC [Str10, p. 5] based on random access stored program (RASP, see [Har71, pp. 234-240]) logic. The operating environment often is a blending of an (typically scaled down) operating system and IT applications known as firmware (see e.g. [Str10, p. 6]). This firmware is often changeable with the means of the system operated by the user (e.g. for product refinement), which makes it potentially vulnerable to malicious firmware updates. Further, e.g. in modern Solid State Drives (SSD), even benign firmware characteristics such as a built-in garbage collection can alter the content of the device (e.g. [BeB10, p. 5]). Here even the powering on of the device can lead to potential evidence being compromised.

In principle, all of the components marked in Figure 45 are susceptible to attacks. To highlight the problem, from literature we exemplarily point towards attack descriptions for internal (SATA, e.g. [Dom19], [RXS+13, pp. 281-286], [Kas19, pp. 1-44]) and external (USB) mass storage (e.g. [NYE17, pp. 675-688]), the network interface card (e.g. [Del20, pp. 1-40] and the BIOS/Firmware [KaK20, pp. 1-81].

For our practical experiments using the Data-Centric Examination Approach DCEA we choose USB attached mass storage on the grounds that the usage of USB thumb drive is widespread and common and we want to highlight potential sources of loss by overlooking maliciously hidden data and uncertainty regarding the identity of devices.

For this scenario we access the mass storage data stream DS_T (see Section 4.2.1) directly and it is required that we have access the chosen component beyond its standard functionality of reading and writing storage blocks of the storage area (USB and SCSI device abstraction, see [Pol20, p. 26]) by issuing commands to the System on Chip (SoC) of the external USB thumb drive, which will also affect the integrated circuits (IC) used for flash memory mass storage.

6.2 Exemplary chosen use case - hidden data in UBS mass storage and device impersonation

USB thumb drives are often part of an examination, examiners present at a suspected incident scene are advised to collect them if visible [ACPO20, p. 12] and to look for their presence in digital traces contained the host IT system (see also [LCL+14, p. 664]). Generally, the USB thumb drive in itself is essentially an IT system with its own mass storage DS_T, main memory DS_M and network DS_N data streams, employing its own processing units (System on Chip SoC, see [Str10, p. 6]) and internal or externally volatile memory (RAM) and non-volatile memory (EEPROM/Flash). It communicates with the host IT system using standardised interfaces, in our use case USB, with standardised command sets for mass storage functionality (e.g. the SCSI command set, see [Cas11, p. 447]). Data on those USB thumb drives can be hidden using a variety of mechanisms (e.g. using a hidden partition patch, user interface tricks, file system specific hiding techniques, etc. see [NYE17, pp. 682-683]).

We concentrate on the option of hiding complete storage sections of the USB thumb drive in a way that it will not be acquired using the block-level data gathering as described in Section 3.3.1.5. The manipulation does not manifest itself in an easily spottable mismatch regarding the device storage capacity. Even leaving the context of the Data-Centric Examination Approach DCEA by including physical inspections, i.e. opening the casing and reading the storage chips label is often unsuccessful. Probably due to economical reasons, the label sometimes is not present (Figure 46).

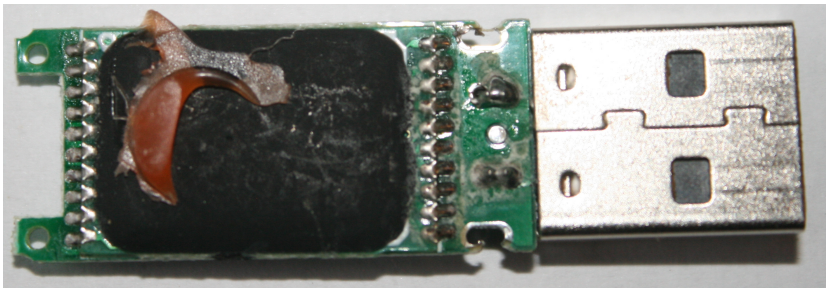

Figure 46: USB thumb drive backplane with no visible clues regarding the storage capacity based on the inscription of the flash memory IC

At least in this example the leads and solder points are accessible, so in a high profile case with the expectation of malicious alterations, an electronics engineer could get access to mass storage flash integrated circuit. If, however, the entire silicon chip is placed directly on the PCB with no intermittent wiring or solder joints, for all but the best-equipped labs with a clean room and specialised desolder equipment, there is little or nothing to be done. But even if a label is present

sometimes the information contained, at least regarding storage capacity derived from the flash memory ICs, this information can be misleading. The storage capacity has been halved from 8GB nominal capacity of the flash memory integrated circuit (IC) M12KX476QH to 4GB (Figure 47).

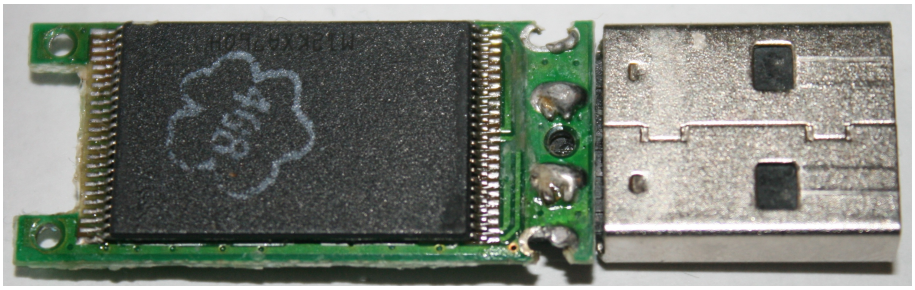

Figure 47: USB thumb drive circuit backplane with flash memory IC M12KX476QH and a stamp claiming half (4GB) of the nominal (8GB) storage capacity

At least at this occasion that reduction has been marked with the "4GB" imprint on the IC housing, probably due to portions of the storage area known to be defective or unreliable during production and thus only the usable portion of the storage area is presented to the host system.

Further, we alter data that is presented as hardware data (e.g. device name at USB and SCSI interface abstraction level, VID, PID, serial number), arguing for the necessity for the level of detail specification (Section 4.4.1.2) when trying to qualitatively determine loss, error and uncertainty using the Data-Centric Examination Approach DCEA introduced in Section 4.

6.3 Practical tests

We employ a common desktop IT system as the host and an USB thumb drive as an example of USB based storage client. With regards to the desktop host, we operate on the mass storage stream DS_T (see Section 4.2.1). The manipulation will not manifest itself in an easily to spot mismatch regarding device storage capacity addressable by the host, thus in details about data DT_3. Further, we even manipulate the device identification on USB and SCSI protocol layer as well as the serial number data detected by the host as hardware data DT_2. We manipulate the USB thumb drive as a storage client on the researchers IT system acting as a desktop IT system host with the aim of hiding data and alter data containing hardware information. We then conduct a forensic examination on the manipulated thumb drive on the examiners IT system acting as a desktop IT system host. Based on the ground truth of the data content regarding the forensic data types DT for digital forensics DF prior to the manipulation, we report deviations from that ground truth. For the ordering of the forensic data types and sets of methods of the forensic process, we again rely on the flow of the sets of examination steps as suggested in Section 4.4.1.

Since the work regarding the practical tests is previously unpublished, we will describe our proceedings here in detail and will continue with the description of the forensic process afterwards. Our experiments (as opposed to [Pol20, pp. 35-39]) do not require new hardware designs or newly created proof of concept software. This work has been done for us by the manufacturers of USB thumb drive designs and are freely available (see e.g. [UDR20]), they are called mass production tools (mp-tools) and are supposed to support the bulk manufacturing process (with regards to firmware) by original equipment manufacturers (OEM) or USB Flash Disk tools (UFD tool). It is at least doubtful, whether those tools are supposed to be available so readily but nonetheless they are. So the complexity shifts from programming a tool (including firmware reversing etc.) to try those existing tools. The usage of those mass production tools can render an USB thumb drive permanently unusable, as we can confirm as it happened many times to us as well. Nonetheless, if a working pair of USB thumb drive and corresponding mass production tool has been identified, all sorts of manipulations (including those explained below) are possible. A great aid in finding a matching pair is the tool ChipGenius [UDR20a] as it gives detailed information about the circuitry inside the thumb drive. To test for the data hiding property we execute the following general test setup:

1. Boot into Windows and inquire the SoC and flash mass storage ICs using ChipGenius Boot into Linux (in our example Ubuntu 18.04.04 LTS) and zero out the USB thumb drive using the ddrescue tool [FSF20] for sanitisation (see Section 3.3.1.6),
2. Boot into Windows and format the drive to a supported file system using normal formatting procedures for the OS (in our example Windows 7 SP1),
3. Collect metadata about the thumb drive (incl. its size) using a trusted forensic tool (box), in our example WinHex [XWS20],
4. Copy a file of a known size first to act as a spacer onto the file system on the USB thumb drive to ensure the hidden data is written past the area to be resized (in our case the file spacer_512.img with a filesize of 512MB),
5. Copy the data to be hidden onto the USB thumb drive (in our case a text file secret.txt with a filesize of 4kB containing the string "this message is secret and shall not be found"),
6. Verify the presence of the data to be hidden using a trusted forensic tool (box), in our example WinHex [XWS20],
7. Resize (shrink) the storage capacity of the thumb drive using the UFDtool (high-level formatting only!) down to the size of the spacer file or below (in our example using UFDUtility_v3.4.8.0 [UDR20b]),
8. Eject/re-insert the thumb drive and zero out the device using the conventional operating system (in our case using ddrescue on the Ubuntu 18.04.04 LTS environment),
9. Format the drive to a supported file system using normal formatting procedures for the OS (in our case in Windows 7 SP1),
10. Check for the presence of the data to be hidden using a trusted forensic tool (box), in our example WinHex [XWS20],
11. In the FUND tool choose the original nominal size and perform a high-level format (in our example using UFDUtility_v3.4.8.0 [UDR20b]),
12. Search for the presence of the hidden data using a trusted forensic tool (box) in our example WinHex [XWS20].

In the following we show an example of this sequence. Our thumb UTD_1 drive is labelled as Intenso Rainbow with a nominative 16GB capacity. It employs a UT165 A1B chipset with the firmware BM 3795 and one Micron MT29F64G08CFABA flash mass storage integrated circuit according to the Chip-Genius_v4_00_1024_0047 tool [UDR20a] as seen in Figure 48.

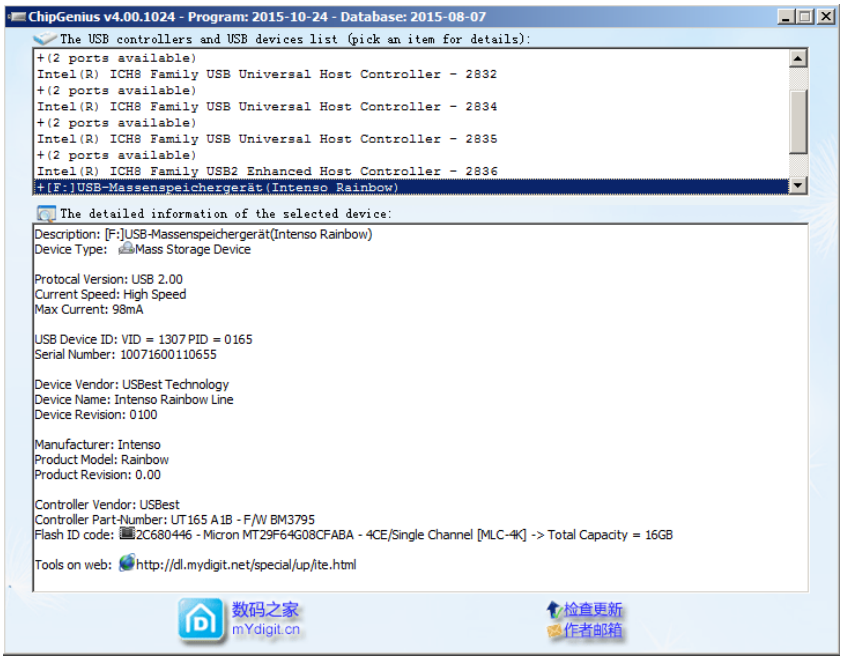

Figure 48: USB thumb drive UTD_1 identification with regards to system on chip SoC and mass storage flash IC (Step1) using ChipGenius_v4_00_1024_0047 [UDR20a]

Following that we boot into Ubuntu 18.04.04 LTS and zero out the USB thumb drive UTD_1 using GNU ddrescue [FSF20] as shown in Figure 48.

```
me@ThinkDifferent:~$ sudo ddrescue --force /dev/zero /dev/sdb
[sudo] password for me:
GNU ddrescue 1.22
     ipos:    15506 MB, non-trimmed:        0 B,  current rate:   1703 kB/s
     opos:    15506 MB, non-scraped:        0 B,  average rate:   3108 kB/s
non-tried:     9223 PB,  bad-sector:        0 B,    error rate:       0 B/s
  rescued:    15506 MB,   bad areas:        0,        run time:  1h 23m  7s
pct rescued:    0.00%, read errors:        0,  remaining time:         n/a
                              time since last successful read:         n/a
Copying non-tried blocks... Pass 1 (forwards)
ddrescue: Write error: No space left on device
me@ThinkDifferent:~$
```

Figure 49: USB thumb drive sanitisation using GNU ddrescue [FSF20]

After booting back into the Windows 7 environment we format the entire storage space of the USB thumb drive UTD_1 into a single FAT32 partition. We collect hardware data about the USB thumb drive UTD_1 using the WinHex [XWS20] forensic tool suite and copy two files, spacer_512.img (512MB in size of arbitrary content) and afterwards secret.txt (4kB on the file system) onto the partition. The sequence is crucial; we want to ensure a certain position of content to be hidden. Using the WinHex forensic tool suite, we verify the presence of the secret file as can be seen in Figure 50.

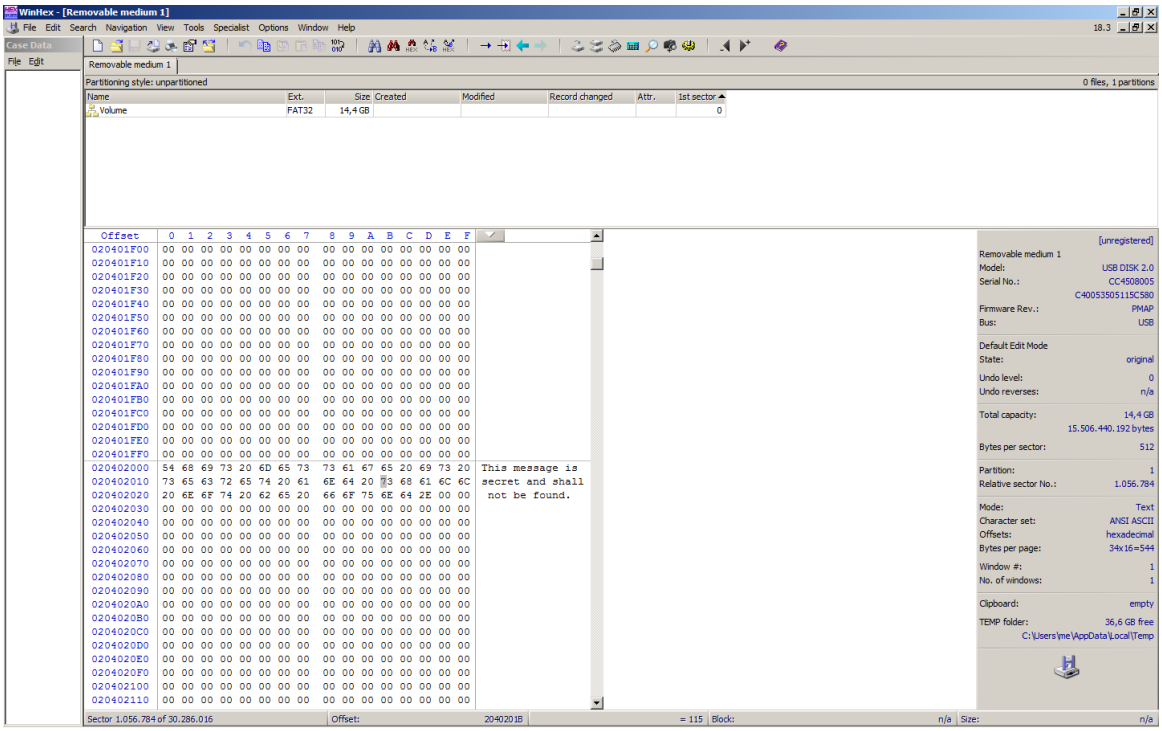

Figure 50: USB thumb drive the USB thumb drive UTD_1 hardware information (left side) and successful search for the text contained in the file secret.txt on the filesystem using the WinHex [XWS20] forensic tool suite (Steps 4 and 7)

Afterwards we use the USB Flash Disk tool UFDUtility_v3.4.8.0 [UDR20b] to shrink the capacity to a size up to or below 512MB spacer_512.img file (Figure 51). This is for demonstration purposes only; in a true data hiding setup one would choose a spacer file that is very close to the nominal capacity of the drive and hide the data in the small margins between the spacer file and the nominal capacity border.

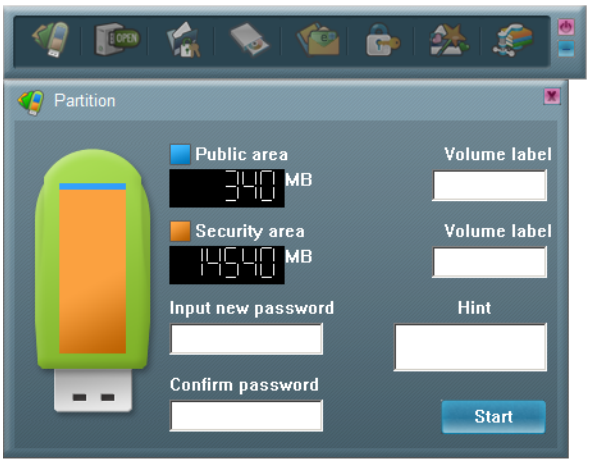

Figure 51: Artificial storage size reduction (step 8) of the USB thumb drive UTD_1 using the USB Flash Disk tool UFDUtility_v3.4.8.0 [UDR20b]

In the following step we reboot into the Linux environment and zero out the size-reduced USB thumb drive UTD_1. Back in the Windows 7 environment we format the flash drive and fail to recover our secret message (Figure 52) using the WinHex forensic tool suite.

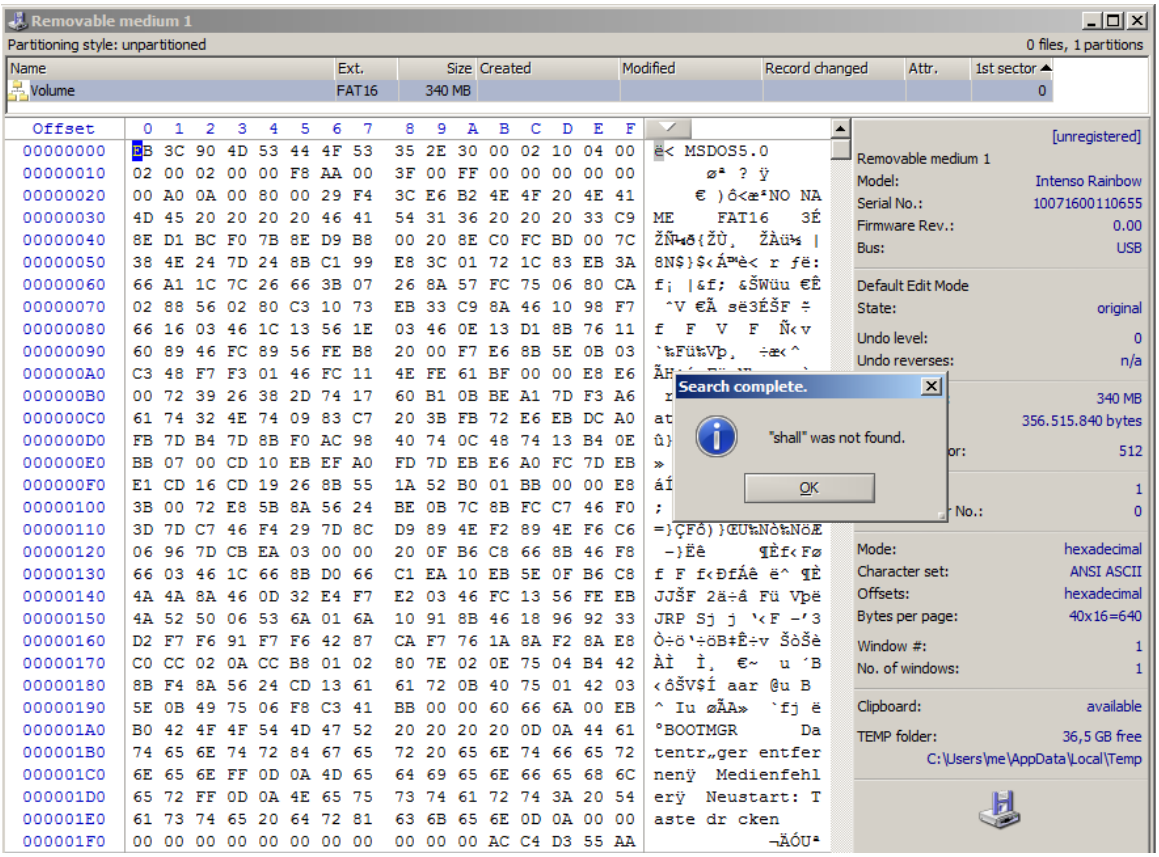

Figure 52: Size-reduced USB thumb drive UTD_1 and failure to recover the hidden string using the WinHex [XWS20] forensic tool suite (Step 11)

Now we resize the USB thumb drive UTD_1 back to its original size and quick format the drive without a zero out procedure. We can now successfully recover the hidden message (Figure 53).

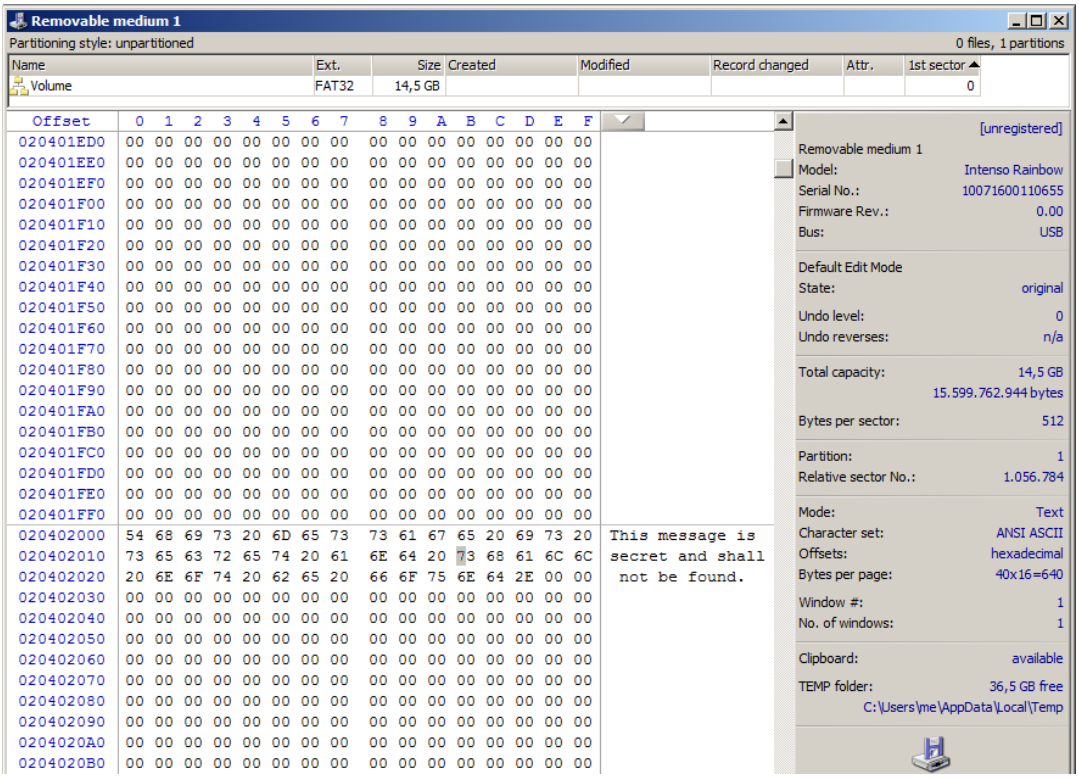

Figure 53: Full sized USB thumb drive UTD_1 and success in recovering the hidden string using the WinHex [XWS20] forensic tool suite (Step 13)

After the data hiding experiment we use a different USB thumb drive UTD_2 for the research on hardware information. Our UTD_2 is, according to its label, a Platinum 16GB USB thumb drive with a nominative 16GB capacity. It employs a Phison PS2251-67 chipset with the firmware 06.05.34 and one Toshiba TC85NVG7T2JTA00 flash mass storage integrated circuit according to the Chip-Genius_v4_00_1024_0047 tool [UDR20a] as seen in Figure 54.

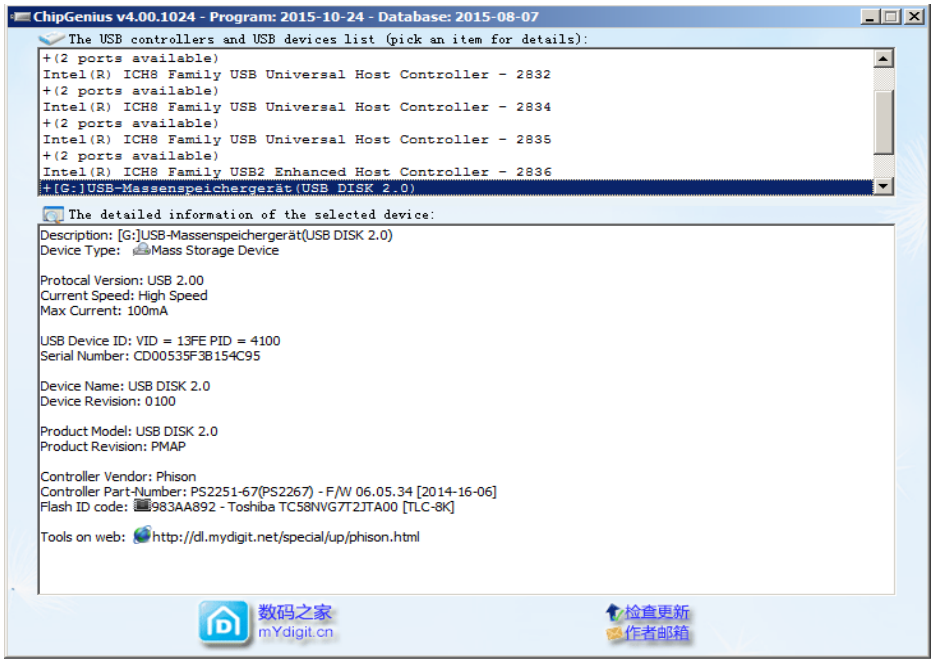

Figure 54: USB thumb drive UTD_2 identification with regards to system on chip SoC and mass storage flash IC using Chip-Genius_v4_00_1024_0047 [UDR20a]

It is recognised in the WinHex forensic tool suite as shown in the following Figure 55.

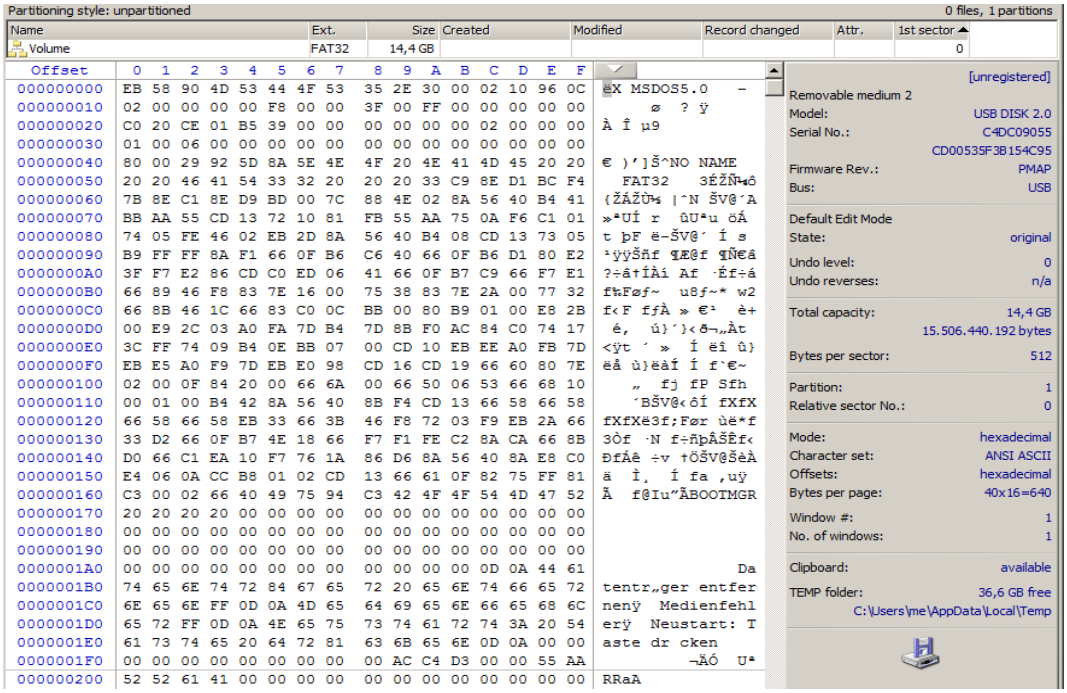

Figure 55: USB thumb drive UTD$_2$ hardware information (left side) using the WinHex [XWS20] forensic tool suite

Note the matching serial numbers and description (USB DISK 2.0). Now we use the mass production tool (mp-tool) UPTool_Ver2093_20150312 [UDR20c] to set the serial number, vendor name, product name and revision for both USB and SCSI interface level to arbitrary values Figure 56.

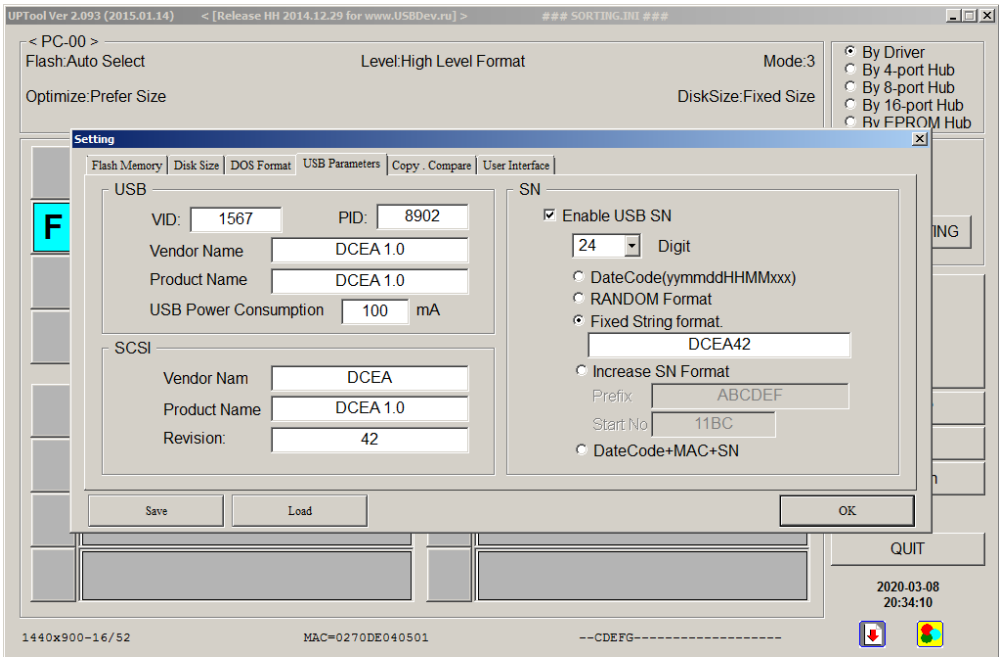

Figure 56: Entering arbitrary values for hardware related information into the mass production tool UPTool_Ver2093_20150312 [UDR20c] for the reprogramming of USB thumb drive UTD$_2$

The changes are reflected in both the ChipGenius tool output Figure 57 and in the output of the WinHex forensic tool suite Figure 58.

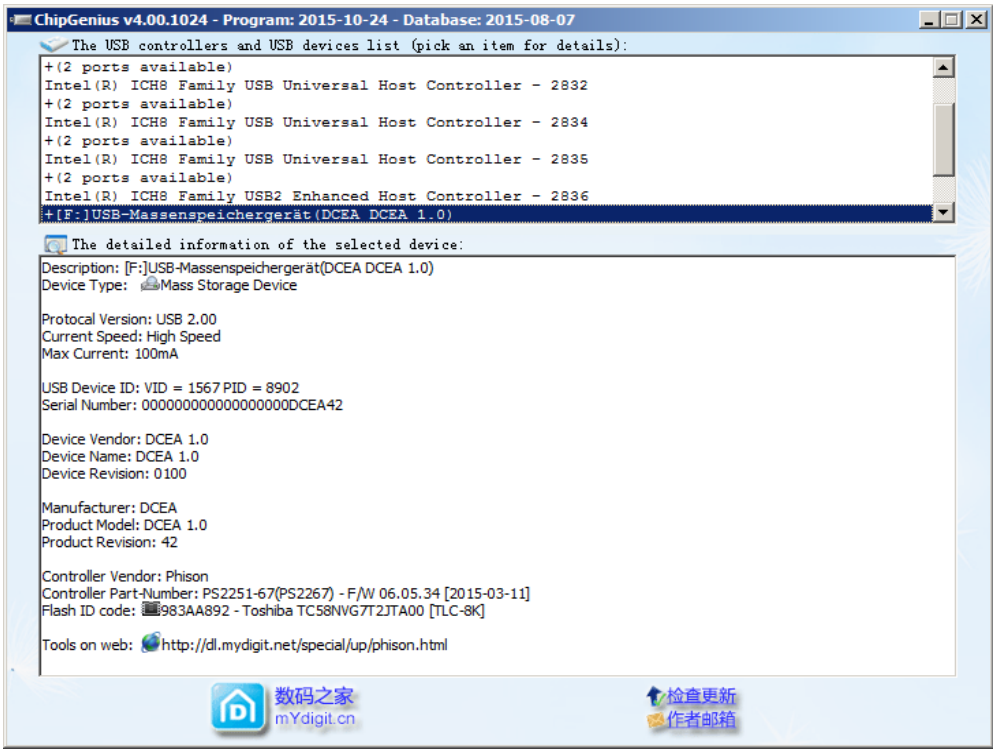

Figure 57: USB thumb drive UTD$_2$ identification with regards to system on chip SoC and mass storage flash IC after the firmware parameter manipulation using Chip-Genius_v4_00_1024_0047 [UDR20a]

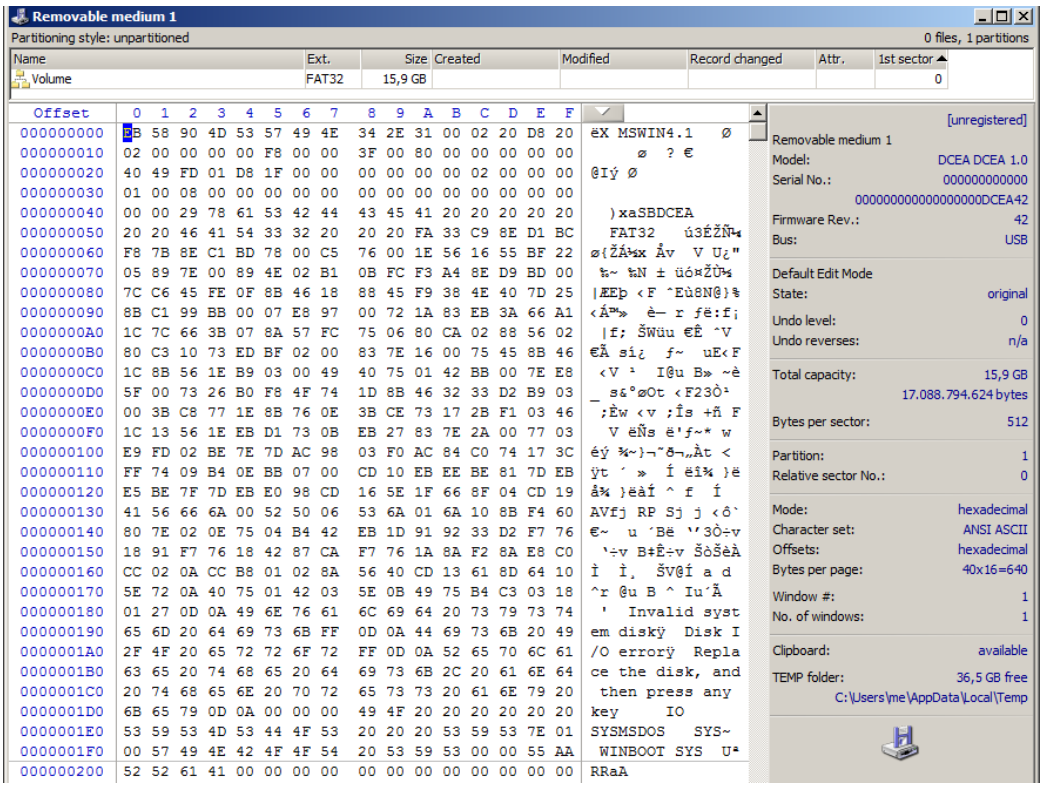

Figure 58: USB thumb drive UTD$_2$ hardware information (left side) using the WinHex [XWS20] forensic tool suite after the firmware parameter manipulation

We will now continue with the description of the forensic process in which we conduct the examination to component-level.

We perform an exemplary evaluation of loss on the forensic data type of user data DT_8 and uncertainty for hardware data DT_2 for digital forensics DF and apply the set relationships regarding this forensic data type using the Venn diagram representation from Sections 4.

6.3.1 Strategic preparation SP_{DF} for the use case of hidden data in USB mass storage and device impersonation

As in the prior use case description in Section 5.3.1 we select items from the description of the strategic preparation SP_{DF} from Section 4.4.1.1 we suggest as being relevant for the purpose of the exemplary examination. Our following description of our findings is only possible due to the strategic preparation SP in particular for the gaining of a ground truth regarding the volume information storage-capacity wise and the hardware information. We start by performing a system landscape analysis as suggested in Section 4.4.1.1 as depicted by the following Figure 59.

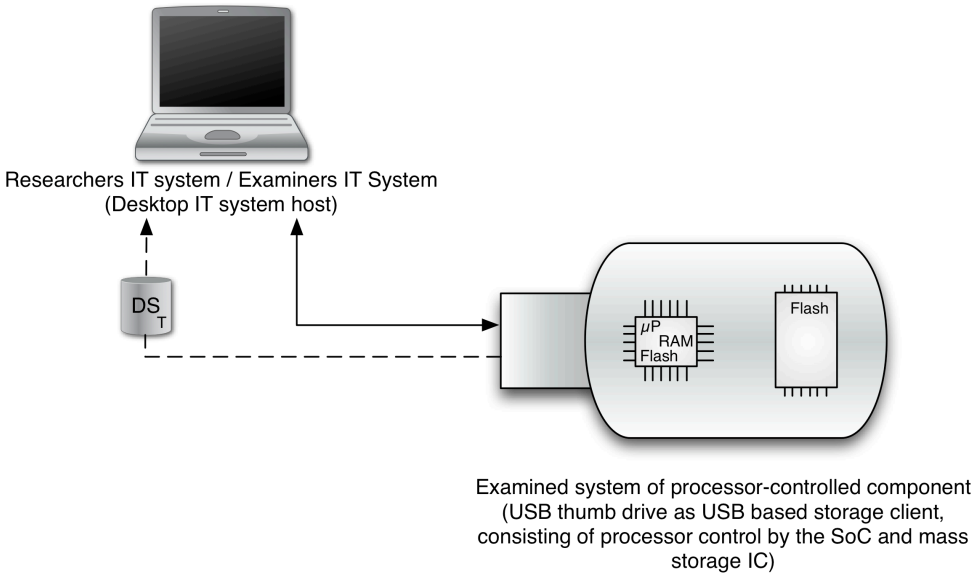

Figure 59: System landscape analysis as part of the strategic preparation SP_{DF} for the use case of hidden data in USB mass storage and device impersonation with the system under examination, the examiners system and the researchers system and the data stream of mass storage DS_T depicted with a dashed line

The researchers system is used to collect the ground truth with regards to storage capacity and hardware information and to manipulate it afterwards in preparation for the following forensic examination. For the forensic examination we use the same IT system as the examiners system on the grounds, that both roles in our case require the same software. The target of the manipulation is the examined IT system in the shape of the USB based storage client, i.e. the USB thumb drive, which consists of the System on Chip (SoC) providing the processor control and the mass storage providing integrated circuit (IC) typically implemented as flash read only memory (ROM). It is directly attached to the researchers IT system using the USB port. We gather, investigate, analyse and document the USB thumb drive using the examiners IT system. We solely rely on the mass storage data stream DS_T. The lines regarding DS_T are dashed in Figure 59 because they do not represent a permanent connection but the origins of data to be investigated, analysed and documented on the examiners IT system. We also prepare an inventory regarding the software installed on the researchers/examiners IT system (Table 71).

Software on the researchers IT system	Exploited sets of methods for the forensic process
Microsoft Windows 7 SP1 Ubuntu 18.04.04 LTS	OS OS
WinHex V18.3	OS DPE
ChipGenius V4 00 1024 0047	ITA
UFDUtility V3.4.8.0	ITA
GNU ddrescue V1.22	OS

Table 71:Software installed on the researchers/examiners IT system as part of strategic preparation in digital forensics SP_{DF}

We look at the methods we use from those installed software items based on the sets of methods for the forensic process in digital forensics DF (Section 4.3.1). The system has a dual boot installation of Windows 7 SP1 and Ubuntu 18.04.04 LTS installed as the operating systems and we will use them as a provider for methods for the forensic process from the operating system OS. For our research purposes we use WinHex forensic tool suite [XWS20] for data gathering, data investigation, data analysis and documentation. As a tool suite it provides numerous sets of methods, we use set of methods for the forensic process of the operating system OS and sets of methods for data processing and evaluation DPE. We further use GNU ddrescue [FSF20] as part of the Ubuntu 18.04.04 LTS Linux distribution, mainly for the provision of sanitisation purposes (see Section 3.3.1.6), which exploits methods for the forensic process of the operating system OS. We further use sets of methods for the forensic process provided by IT application ITA of the ChipGenius tool [UDR20a], which displays various hardware information for USB-based mass storage devices and of the mass production tool UFDUtility_v3.4.8.0 [UDR20b], which allows us to set/read mass storage configurations and hardware information for USB-based devices.

Note that the mass production tool UPTool_Ver2093_20150312 [UDR20c] does not get a mention because it can only change data but not inquire them and are thus not helpful in the proceedings of the forensic examination. It is, however, instrumental in the research and in the conclusion in Section 6.4.

All of those case-independent actions are documented as part of the process accompanying documentation as part of the documentation DO_{DF} step (see Section 6.3.6). With those preparations we are sorted for the invocation of a forensic examination.

6.3.2 Operational preparation OP_{DF} for the use case of hidden data in USB mass storage and device impersonation

During the examination step of operational preparation OP_{DF} we provide a storage folder on the examiners IT system initiate a chain of custody for digital objects by the provision of a structured and protected storage space on the mass storage of the examiners system. We are supported by the built-in case-management and reporting functionality of the purpose-built forensic tool suite WinHex [XWS20]. We deem the examiners IT system fit for purpose based on the software installed according to Table 71 from Section 6.3.1.

For our research involving a forensic examination we set the level of detail (see Section 4.4.1.2) to component level, which means we examine processor-controlled components (in our case the USB thumb drive) separately. We assume that the experiments described in Section 6.3 regarding the data hiding approach involving USB thumb drive UTD_1 and the hardware data manipulation involving USB thumb drive UTD_2 are conducted and look for conventional means to detect those alterations and see how alternative, currently untested methods can improve results.

All of those case-specific actions are documented as part of the process accompanying documentation DO_{DF} step (see Section 6.3.6).

6.3.3 Data gathering DG_{DF} for the use case of hidden data in USB mass storage and device impersonation

We summarise the step of data gathering for digital forensics DG_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 72:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
WinHex: Create Disk Image	DS_T	DT_1	OS	DT_1, DT_3	D_{IF}, D_{IFC}
WinHex: Open Disk	DS_T	DT_1	OS	DT_2, DT_3, DT_4	D_{IF}, D_{IFC}
ChipGenius	DS_T	DT_1	OS	DT_2,	D_{IF}, D_{IFC}
UFDUtility: Device manager	DS_T	DT_1	OS	DT_4	D_{IF}, D_{IFC}

Table 72: Summary of the actions taken during data gathering DG_{DF} based on [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

During data gathering DG_{DF} we try to access the data in its lowest and most encapsulated form (see also Section) and thus try to access our data from the investigation target D_{IT} at the lowest and most comprehensive way by exploiting methods of the forensic process offered by the operating system OS of the USB thumb drive. However, we should not forget that we are getting a filtered view due to the processor-controlled access to the mass storage flash integrated circuits. The following statements will apply for both the USB thumb drive UTD_1 and UTD_2 from the experiments conducted in Section 6.3. At this early stage, we expect both data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC} to be gathered.

This system on chip SoC (see Figure 59 depicting the system landscape analysis) will report the number of mass storage sectors and grant access to them.

The data of the investigation result D_{IR} using the method Create Disk Image provided by the forensic tool suite WinHex [XWS20] exploiting the operating system of the USB thumb drive is raw data DT_1 off the mass storage data stream DS_T provided by the USB thumb drive with regards to the mass storage data content and associated meta data (e.g. file size of the forensic image) as details about data DT_3. We disregard DT_3 for our experiments but document their presence during process accompanying documentation.

Further, the method Open Disk from the WinHex forensic tool suite accesses raw data DT_1 from the USB thumb drive using the USB thumb drive's operating system and provides without further interaction a summary of selected data (see e.g. Figure 55) such as serial number, firmware revision etc. representing hardware information DT_2, partitioning representing details about data DT_3 and size information (total capacity, bytes per sector etc.) which are configurable and thus represent configuration data DT_4.

The ChipGenius software (see also Figure 48) accesses the raw data DT_1 provided by the methods of the operating system OS from the USB thumb drive and gathers hardware data DT_2 (including serial number, SoC, flash IC etc.).

The UFDUtility also accesses the raw data DT_1 provided by the methods of the operating system OS from the USB thumb drive but gathers configuration data DT_4 in the shape of the public area of the mass storage flash IC content (see also Figure 51). Some of the data gathered will be directly picked up during data analysis DA_{DF} examiner in Section 6.3.5.

Strictly speaking, neither ChipGenius nor UFDUtility represent forensic software with the latter also having the option of write access to the USB thumb drive (which we will exploit in an experimental manner in the step of data analysis DA_{DF} in Section 6.3.5).

We perform the data gathering for the USB thumb drive UTD_1 twice in search for the hidden data and for the device impersonation on the USB thumb drive UTD_2 based on the step-by-step experiment described in Section 6.3 and initiated by our findings from the data analysis step DA_{DF} (see Section 6.3.5). Both those re-runs are perfectly in tune with our description of the flow of examination steps allowing backward steps (see Section 4.4.1.7).

All of those case-specific actions are documented as part of the process accompanying documentation DO_{DF} step (see Section 6.3.6).

6.3.4 Data investigation DI_{DF} for the use case of hidden data in USB mass storage and device impersonation

We summarise the step of data investigation DI_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 73:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
WinHex: Technical Details Report	DS_T	DT_2, DT_3, DT_4	DPE	DT_2, DT_3, DT_4	D_{IF}, D_{IFC}
WinHex: Find Text	DS_T	DT_1	DPE	DT_8*	D_{IF}, D_{IFC}

Table 73: Summary of the actions taken during data investigation DI_{DF} based on [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained, * data only available in the second run after the modification of the USB thumb drive UTD_1

In the examination step of data investigation DI_{DF} we operate on the data gathered in the previous step in an effort to reduce the amount of data that is data containing forensically relevant information D_{IF} but not also part of data containing case-specific, forensically relevant information D_{IFC}. Our methods for the forensic process used in this step and in this use case require a modest amount of user interaction but execute fairly autonomous.

For the investigation into hidden data involving the USB thumb drive UTD_1 we use the Find Text method of the forensic process provided by the forensic tool suite WinHex [XWS20]. It operates on raw data DT_1 of the mass storage data stream DS_T acquired during data gathering DG_{DF} form-

ing the data of the investigation target D_{IT} and is a method from the sets of methods for data processing and evaluation DPE. In our experiment we look for user data DT_8 as text content as part of our search for the hidden text file as the data of the investigation result D_{IR}.

In the first run, involving only commonly accepted methods for the forensic process for mass storage examination we do not find the particular text string "shall" from our known hidden text file secret.txt (see Section 6.3). In a second run (which involves another data gathering step DG_{DF} and another data investigation after finishing data analysis DA_{DF} (see Section 6.3.5), which is perfectly in tune with our description of the flow of examination steps allowing backward steps (see Section 4.4.1.7), we do find the text string and indeed the whole text (see Figure 53).

For the investigation into technical details regarding the USB thumb drive UTD_2 we use the Technical Details Report method of the forensic process provided by the forensic tool suite WinHex [XWS20]. It operates on raw data DT_1 of the mass storage data stream DS_T acquired during data gathering DG_{DF} forming the data of the investigation target D_{IT} and is a method from the sets of methods for data processing and evaluation DPE. It returns hardware data DT_2, details about data DT_3 and configuration data DT_4 as data of the investigation result D_{IT}. We disregard DT_3 for our experiments but document their presence at process accompanying documentation.

In the first run it returns the initial hardware data as well the details about data and configuration data, which we generally leave unaltered. We alter the hardware data DT_2 as described in Section 6.3 and re-run the data gathering DG_{DF} from Section 6.3.3 prior to another data investigation step DI_{DF} for UTD_2.

All of those case-specific actions are documented as part of the process accompanying documentation DO_{DF} step (see Section 6.3.6).

6.3.5 Data analysis DA_{DF} for the use case of hidden data in USB mass storage and device impersonation

We summarise the step of data analysis in digital forensics DA_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 74:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
UFDUtil: Partition manager	DS_T	DT_4	OS	DT_4	D_{IFC}
Human examiner	DS_T	DT_2	DPE	DT_2	D_{IF}, D_{IFC}

Table 74: Summary of the actions taken during data analysis DA_{DF} based on [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

In the examination step of data analysis DA_{DF} we use the data from the step of data gathering DG_{DF} directly and from the previous step of data investigation DI_{DF} in an effort to further reduce the amount of data containing forensically relevant information D_{IF} but not data containing case-specific, forensically relevant data D_{IFC}.

This step relies heavily on prior case-specific knowledge from the previous steps and from the experiments staged on the same IT system shared by experimenter and examiner. DA_{DF} in this experiment is only assisted by the methods described in the following. We highlight this by adding the methods for the forensic process provided by the examiner as sets of methods for data processing and evaluation DPE (see Table 74). This is due to the experimental nature of the examination of the use case as research work. We rely on backward steps to both data gathering DG_{DF} and data investigation DI_{DF}.

For the investigation into hidden data involving the USB thumb drive UTD_1 we use the partition manager method UFDUtil software, which exploits a method of the operating system OS of the USB thumb drive. As data of the investigation target D_{IT} it queries the configuration data DT_4 concerning the memory layout and boundaries of the flash memory mass storage integrated circuit (IC). We use the method to display/visualise those data as configuration data DT_4 to the forensic examiner in the first run.

We describe the human examiner's action as a forensic method. For the actions on hidden data involving the USB thumb drive UTD_1 the forensic examiner acts on the presence of the data type of user data DT_8 and the configuration data DT_4 of the thumb drive UTD1 as data of the investigation target D_{IT}. The actions of the human examiner represent methods of data processing and evaluation DPE from the set of methods for the forensic process. The data of the investigation result D_{IR} are user data DT_8 and configuration data DT_4.

As an artificial size reduction is noticed (see Figure 51), the examiner decides to deliberately alter the configuration to restore the storage area of the USB thumb drive whilst reducing the data interference/integrity violation to a minimum by avoiding any low level flash memory initialisation/formatting. This conforms to the careful weighting process described in Section 3.3.1.3. However, a high-level formatting resulting in a newly initialised file allocation table does occur, meaning that for a full reconstruction, the undelete mechanisms described in Section 4.2.1.1.2 need to be applied. We restrict ourselves to only find the traces of the file and thus disregard this further effort. In a second run, using the altered USB thumb drive UTD_1, the whole process starting with data gathering DG_{DF} is re-executed.

In our case the absence of DT_8 can serve as a first qualitative indicator for *loss* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA, where our ground truth describes the presence of that data we placed there ourselves during the experiment (see also Figure 60 for the Venn diagram representation).

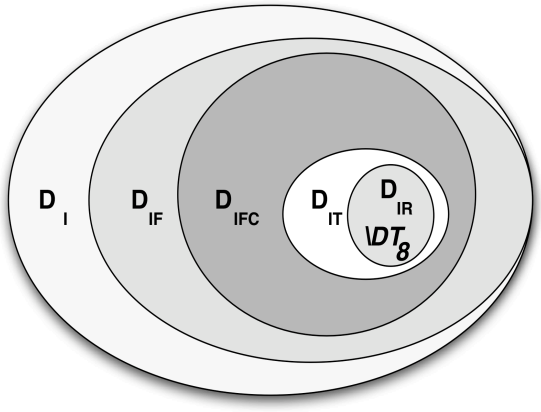

Figure 60: Loss due to the absence of DT_8 (denoted with the "\" symbol) as data of the investigation result D_{IR}, which directly conflicts with the ground truth from our experiments

For the detection of hardware data DT_2 alterations for device impersonation as performed on the USB thumb drive UTD_2, the human examiner looks for potential signs of manipulations by means of detecting mismatches between the outputs of the hardware data DT_2 of the method Technical Details Report of the WinHex forensic tool suite [XWS20] from data investigation DI_{DF} and that

of the ChipGenius method from data gathering DG_{DF}. Here too, we describe the human examiner's actions as methods of data processing and evaluation DPE from the sets of methods for the forensic process with hardware data DT_2 and configuration data DT_4 as data of the investigation target D_{IT}. Data of the investigation result D_{IR} are their evaluation and thus hardware data DT_2 and configuration data DT_4.

Unfortunately, no signs of a manipulation are detectable as of now. However, our experiments reveal the presence of two different contents of the hardware data DT_2. This can serve as a first qualitative indicator for *uncertainty* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA, where our ground truth describes the presence of that data we placed there ourselves during the experiment (see Figure 61 for the Venn diagram representation).

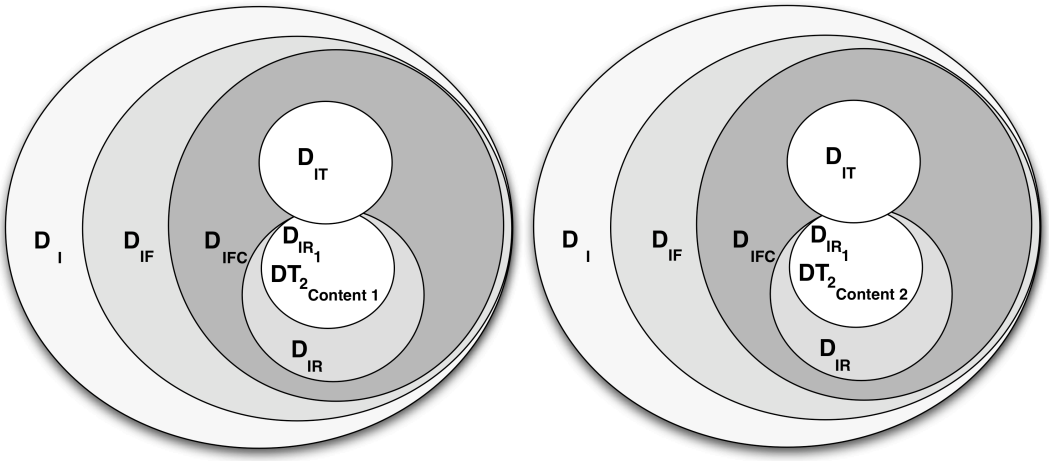

Figure 61: Uncertainty due to content differences (denoted as content 1 and content 2) of DT_2 as data of the investigation result D_{IR} as inflicted during our experiments

All of those case-specific actions are documented as part of the process accompanying documentation DO_{DF} step (see Section 6.3.6). This entails the judgment and the factors leading to the deliberate alteration of the USB thumb drive UTD_1 configuration data DT_4.

6.3.6 Documentation DO_{DF} for the use case of hidden data in USB mass storage and device impersonation

We summarise the step of documentation in digital forensics DO_{DF} as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DT of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DT of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 75:

Specific instantiation of the method of the forensic process	Accessed data streams DS	Forensic data type for data of the investigation target D_{IT}	Exploited method of the forensic process	Forensic data type for data of the investigation result D_{IR}	Involved sets regarding the information contained
WinHex: Compute hash	DS_T	DT_1, DT_2, DT_3, DT_4, DT_8*	DPE	DT_3	D_{IF}, D_{IFC}
WinHex: Log general Activity	DS_T	DT_1, DT_2, DT_3, DT_4, DT_8*	DPE	DT_3, DT_7	D_{IF}, D_{IFC}
WinHex: Include screenshots in log	DS_T	DT_1, DT_2, DT_3, DT_4, DT_8*	DPE	DT_3, DT_7	D_{IF}, D_{IFC}

Table 75: Summary of the actions taken during documentation DO_{DF} based on [KHD+09, pp. 1619-1633] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained, * data only available the second run after the modification of the USB thumb drive UTD_2

As stated in Section 4.4.1.6 and previously practically applied in Section 5.3.6 during the process accompanying documentation as part of the step of documentation in digital forensics DO_{DF}, every step taken, every input data and output data needs to be meticulously and comprehensively documented. The WinHex tool suite offers a number of methods to support the examiner by providing (semi-) automated means do document all actions taken in the confines of the tool suite. For descriptive purposes in this thesis we pick exemplary selections of those methods to show the application of the Data-Centric Examination Approach DCEA. The data gathered and investigated, in our example raw data DT_1, hardware data DT_2, details about data DT_3, configuration data DT_4, user data DT_8, form data of the investigation targets D_{IT} for all exemplary chosen methods provided by the WinHex forensic tool suite. All methods return meta data as details about data DT_3 with regards to the data of the investigation result D_{IR}. Additionally, both methods Log general activity and Include screenshots in logs, also create session data DT_7.

For all intents and purposes of this thesis, this description of proceedings of the whole Section 6 constitutes a final report as part of the examination step of documentation DO_{DF} as described in Section 4.4.1.6.

The highly automated process accompanying documentation built into the WinHex [XWS20] does the entire recording and reporting its commercially available variant X-Ways Forensics [XWS20a] provides. However, the freely available WinHex does not show those reports but generates and keeps them internally. We chose WinHex over its commercially available counterparts on the ground of using the fully functionality of the forensics engine and still providing means for others to re-create and evaluate our findings. Our focus in this thesis is the description of our functional findings and the application of our Data-Centric Examination Approach DCEA. This highlights the experimental character of our research. We are fully aware that for a forensic examination in regular proceedings, the full reporting capabilities are essential. The good thing is that we could load our case files into X-Ways Forensics and would be able to create those reports, which we omitted on the grounds of availability of that program to us.

6.4 Summary of new findings and evaluation with regards to loss, error and uncertainty for processor-controlled components in the use case of hidden data in USB mass storage and device impersonalisation

In summary we can constitute that the examination of the use case can be described using our forensic data types DT for digital forensics DF from Section 4.2.1, the sets of methods for the forensic process in digital forensics DF from Section 4.3.1 and the sets of examination steps for digital forensics DF from Section 4.4.1. We can use the separation of data of the investigation target D_{IT} and the data of the investigation result D_{IR} and the set relations between data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}. This description provides *common language* to comparably describe different forensic examinations in a structured way, thereby allowing to systematically address issues of loss, error and uncertainty within procedures. The following Figure 62 summarises the examination.

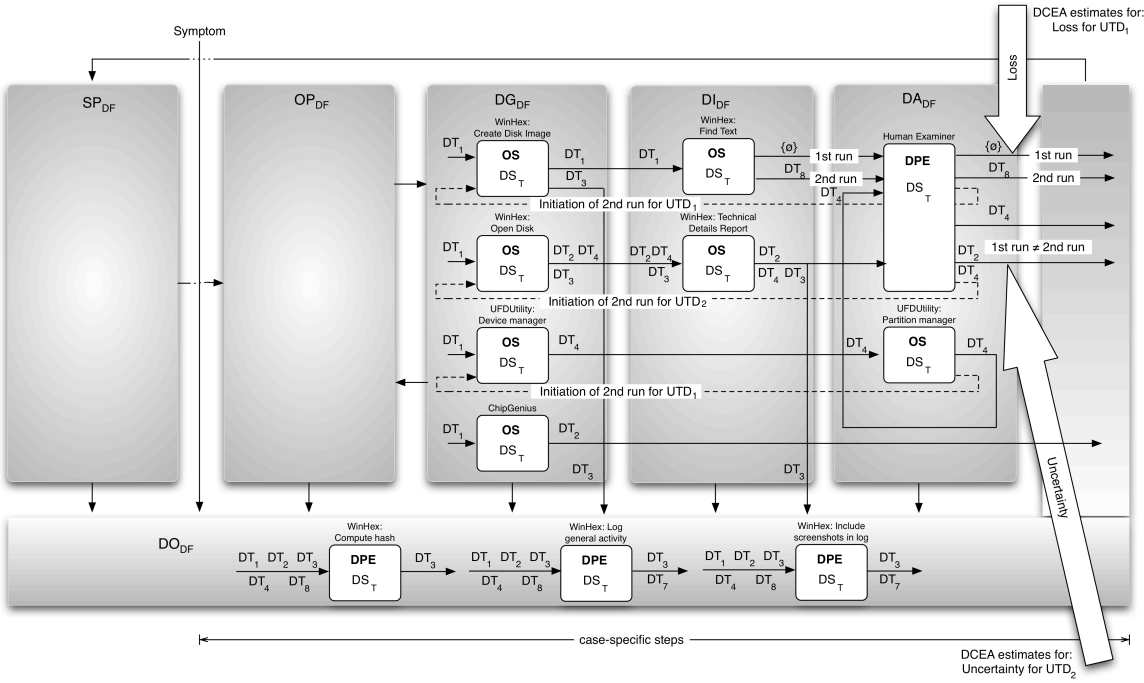

Figure 62: Depiction of the examination flow as summary of the use case of hidden data in USB storage and device impersonation with the involved forensic data types and sets of methods for the forensic process

When describing the use case of hidden data in USB mass storage and device impersonation using the Data-Centric Examination Approach DCEA, we can also use the data types for a first indicator for loss, error and uncertainty in a qualitative estimation as suggested in Section 4.5.

For our experiments on hidden data using the USB thumb drive UTD_1 we can constitute a *loss* for the first run, i.e. without direct alteration of the drive itself. In the first run using conventional, universally agreed mass storage forensics we failed to gather, investigate, analyse and document user data DT_8 (in our example the content of the text file “secret.txt”). We could only recover it in the second run using experimental, closed source software of dubious sources that are badly documented (if at all) that seem to have been helpful in our case but its use is likely to conflict with the requirements regarding the handling for forensic evidence (see also our additional reflections in the appendix Section 10.1.3. Our findings highlight the need for forensic tools instead of highly experimental, closed source applications.

Our experiments regarding the device impersonation using the USB thumb drive UTD_2 are even more severe. They point to an, as yet undiscoverable, *uncertainty* regarding the hardware data presented by processor-controlled components. In other application areas, such as the automotive

industry, the problem of alteration of flash memory content (e.g. Chip Tuning) is discussed (e.g. [HKD08, pp. 245-246]) and a few solutions have been proposed. But, in forensics the examiner is expected to deal with all sorts of environments, often involving legacy components and devices, and the challenged by processor-controlled components seems here to stay.

In general our findings underline the demand for a level of detail selection (see Section 4.4.1.2) and its justification, and the alignment of the expectations with regards to loss, error and uncertainty.

Further, the trust placed on data providing hardware information that was gained from processor-controlled components with possible and easy alterations in memory locations (volatile and non-volatile) should be adjusted accordingly. As stated in [Cas11, p. 488], hardware information on its own should only be used for keeping track of items in case documentation.

7. Application of the approach to digitised forensic dactyloscopy

The contents of this section have been peer reviewed and published in part within the scope of the following publications as joint work with the co-authors Jana Dittmann, Mario Hildebrandt, Marcus Leich, Claus Vielhauer, Michael Ulrich, Ina Großmann (in descending order): [HKG+11], [HML+11], [KLD+11], [KHD12], [HKD+14]

For this use case we use the approach described in [HKD+14, pp. 902808 - 902808-15] regarding digitised forensics, i.e. the digitalisation of physical trace evidence and the digital examination in support of the forensic expert (see Section 2.1.1). Forensic dactyloscopy denotes a forensic discipline focusing on the comparative analysis and evaluation of fingermarks and fingerprints for individualisation purposes [Meu15, p. 734]. We pick exemplary steps as depicted in Figure 63 of the full process chain from [HKG+11, p. 6]. Generally, the approach in [HKD+14, pp. 902808 - 902808-15] is a proof of concept based on ongoing research based on devices, firm/software developed for other purposes for the research into (partial) implementation of the biometric/pattern recognition pipeline (see Section 3.4). We do not know of any ready-made solution for the contactless examination of latent fingerprints. We use this use case to show how the application of the Data-Centric Examination Approach DCEA introduced in Section 4 can give us a qualitative estimation of loss, error and uncertainty and a means to describe the forensic process where at the time of writing the original article no universally agreed procedures existed. This is in effect to answer our research question from Section 1.1:

> Can a data-centric approach be designed to preserve data/tool sovereignty of the forensic examiner and to prevent bias from tool usage result and to reduce loss, error and uncertainty?

on the grounds that we can abstractly describe and judge a first research approach into digitised forensic dactyloscopy.

7.1 Special requirements and properties of digitised forensics

A dominant property shared in many application areas of digitised forensics is the focus on latent traces, i.e. traces that are invisible to the naked eye (see also Section 3.4). This requires a development (see [Bar11, p. 1-13]), i.e. measures to enhance the contrast between the substrate and the latent trace it contains. The major advantage of digitised forensics is the contact-less acquisition of aspects of the trace with sensors and subsequent digitalisation and thus digital-only development of the trace. This, however, requires strict transparency and comprehensibility, more likely even higher than in traditional crime scene forensics. Here we see the Data-Centric Examination Approach DCEA introduced in this thesis as a means to support transparency as it can describe the route the digital data took from acquisition to the final report.

Although biometric systems for user authentication (see Section 2.8) share a lot of properties with the process described below, the main difference is the type of fingerprint we use as a physical basis for the digitalisation. In biometrics we typically use patent (i.e. non-hidden, obvious) features of the consenting person to be identified or verified taken directly off the person. In digitised forensics, typically latent residue and similar features unwillingly left by people at the crime scene are examined as traces. Those traces that can be connected to particular people are typically of a far less quality with regards to their features and are often only partially available. The process leading to the deposition of the traces is not repeatable (in the contrary, the person committing a crime is highly interested to leave as little traces as possible), whereas in biometric systems the user could be asked to repeat the process in case of an unwanted result.

We apply the statement from [InR01, p. 92] whereby the number of things that can be evidence is limited to the things that exist in the physical universe - in other words, anything can be evidence. This way, we only expect data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}.

Since many approaches utilise machine learning, an important requirement is the strategic preparation SP_{SF} for digitised forensics SF, wherein the models are trained. These models are not case-specific but specific to the respective application environment.

Secondly, the results of the sensors often depend very highly on the substrate that contains the suspected trace and the settings, with which the data is gathered, investigated and analysed. The determination of both the best settings and the best sensor for a given substrate occurs not case-specific as part of benchmarking (see Section 4.4.1.1 and e.g. [HML+11, pp. 1-6] and its extension in [KLD+11, pp. 78810G - 78810G-15] and [KHD+12, pp. 1504-1508]) during strategic preparation SP_{SF}.

We depend on the data streams $DS_{1..n}$ delivered from the sensors S_1 to S_n to capture digital representations of the physical world, namely the substrate suspected to carry latent traces (see Section 4.2.2).

7.2 Exemplary chosen use case - Non-destructive latent fingerprint examination

To show the applicability of our Data-Centric Examination Approach DCEA from Section 4 we choose our research into non-destructive latent fingerprint examination based on [HKD+14, pp 902808 - 902808-15] as part of digital dactyloscopy. We use a surface measurement device FRT CWL 600 [FRT20, p. 2] for the acquisition of the substrates containing latent fingerprint traces. It has the capability do digitise planar surfaces and return intensity and topographical data of the acquired surface area, providing us with the sensor data stream DS_{S1}. The chromatic white light sensor device itself is attached to a desktop IT based measurement support system using a USB connection (see also Figure 64 about the system landscape analysis during strategic preparation SP_{SF}). For evaluation purposes of this ongoing research work, we also employ a Smiths Heimann LS1 LITE Xe [FCC+20] live scanner, providing us with the sensor data stream DS_{S2}. It reads exemplary biometric fingerprint data directly off the test subject's finger (live scanner) and returns a black and white image containing fingerprint information. This sensor, too, is attached a desktop IT based measurement support system but using the IEEE 1394 bus instead. Since the research operates on highly data protection sensitive, life-long persistent characteristics of the fingerprints (morphology, i.e. the ridge lines pattern, see [Van11, p. 9-7]), the gathered data is stored in an isolated network consisting of a file server acting as primary storage and the workstations for research work acting as secondary storage. We are entirely bound to meet the data protection requirements (see also additional details in the appendix in Section 10.1.4). For our research we modify the directional flow of examination steps from [HKG+11, p. 6] in that we are only concerned with the detailed scan of the surfaces containing the fingerprints (Figure 63). This is based on the research character of our work, where we exactly know where we placed those latent fingerprint traces. We use the exemplary fingerprint data and the NBIS [NIS+20] biometric suite as a means to deliver a sort of ground truth for our research.

Since our flow describes the whole research (also in support of benchmarking, see [HML+11, pp. 1-6]), we also added our evaluation; including the live biometric scans, as a horizontal plane into the original depiction (dotted line) and all the work necessary during strategic preparation but entails steps also used in the case-specific part of the examination (dashed line). Our devices and, indeed, the software are not purpose-built for forensics but have characteristics (resolution, settings recording etc.) that enable fundamental research into the topic.

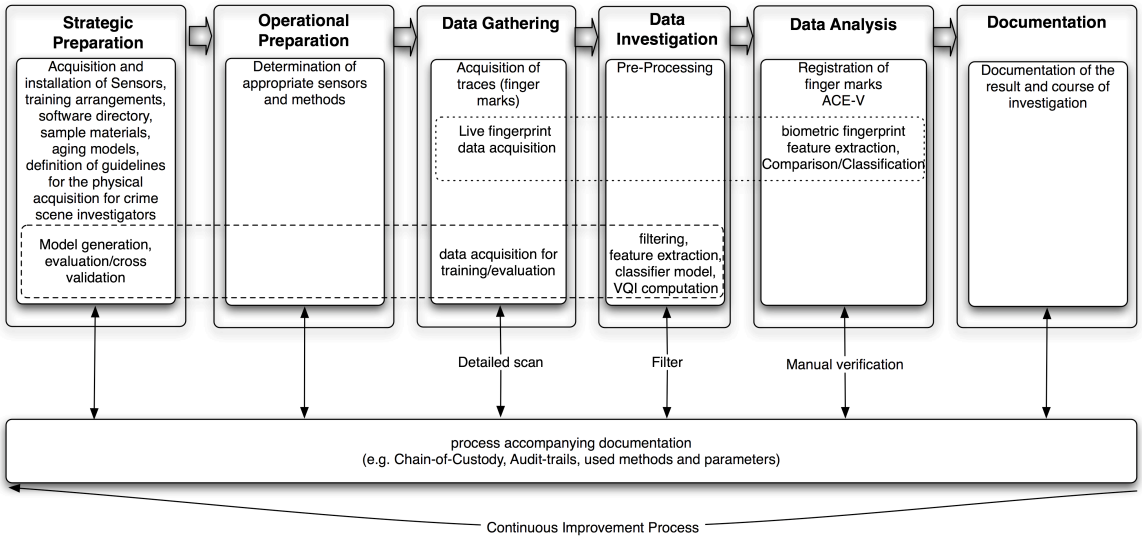

Figure 63: Directional flow of examination steps in digitised forensics on the example of non-destructive latent fingerprint examination based on [HKG+11, p. 6] but exclusively focusing on the digital data flow (omitting physical acquisition), on two sensors (2nd live scan only for evaluation), on the detailed scan during data gathering (omitting profile scan and coarse scan) and on pre-processing during data investigation (omitting separation of overlapping scans and age detection) with added extra items from [HKD+14, p. 902808-6] for model generation and evaluation (dashed line) and for verification against a live scan (dotted line)

7.3 Practical tests

We use the sets of examination steps for digitised forensics SF from Section 4.4.2.1 as a means to order the flow of the examination of data types for digitised forensics DD (see Section 4.2.2), which are processed using the set of methods for the forensic process in digitised forensics SF (see Section 4.3.2) to process the data. This is analogous to the general flow of proceedings as described in Section 4.4.2.7. The following description is a sort of high-level view on our experiments to show the applicability of the Data-Centric Examination Approach DCEA introduced in Section 4. For a more detailed technical description of our experiments we refer the interested reader to the original article [HKD+14, pp. 902808 - 902808-15].

We perform an exemplary evaluation of error and loss on the forensic data type of trace-specific feature data DD_5 for digital forensics SF and apply the set relationships regarding this forensic data type using the Venn diagram representation from Sections 4.

7.3.1 Strategic preparation SP_{SF} for the use case of non-destructive latent fingerprint examination

We start the description of the strategic preparation SP_{SF}, selecting items from the general description in Section 4.4.2.1, analogous to the other use cases with the system landscape analysis (Figure 64).

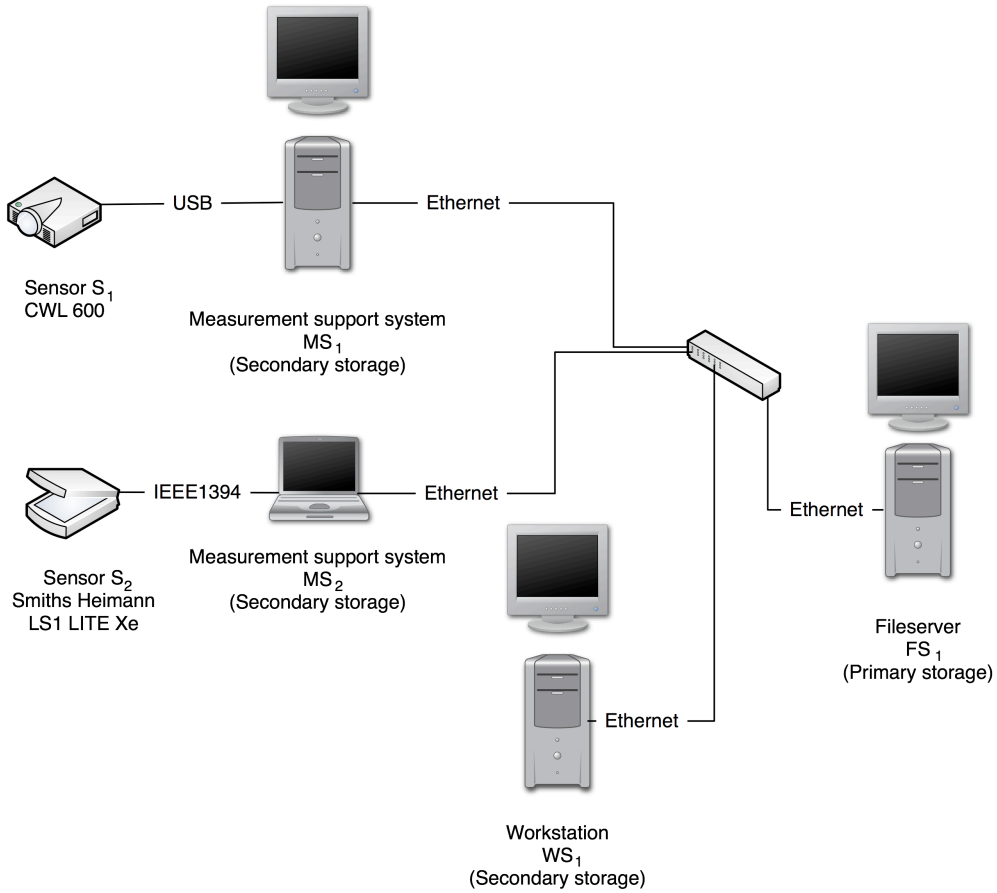

Figure 64: System landscape analysis for the use case of non-destructive latent fingerprint examination with the sensors S_1 and S_2 connected to the measurement support systems (MS_1, MS_2) using bus systems (secondary storage for biometric data) and to the fileserver FS_1 (primary storage for biometric data) using an Ethernet connection that is physically separated from the Internet

Using the Sensor S_1 as provider of the data stream DS_{S1} containing substrate data with latent fingerprint traces, we store the acquired data trough the usage of the connected measurement support system MS_1 on our fileserver FS_1 acting as primary storage. Analogous, the measurement support system MS_2 receives the data stream DS_{S2} from the sensor S_2 containing the live, exemplary scan data and transfers it to the fileserver FS_1. Both MS_1 and MS_2 act as secondary storage, where data is also deleted on demand due to storage capacity limits. The research is carried out on the workstation WS_1. We prepare the software inventory for the measurement support systems MS_1 and MS_2, the fileserver FS_1 and the Workstation WS_1 as shown in Table 76.

Software on the measurement support system MS_1	**Provided sets of methods for the forensic process**
DD+Acquire + FRT Acquire V1.46	DPO, DCI
sha256sum (GNU coreutils 6.11)	CC
Software on the measurement support system MS_2	**Provided sets of methods for the forensic process**
Smiths Heimann TestWizard V 1.10	DPO
sha256sum (GNU coreutils 6.11)	CC
Software on the workstation WS_1	**Provided sets of methods for the forensic process**
Weka V. 3.6.6 implementing SMO, J48, Bagging	MG
GIMP V 2.6.12	IE
QTFPX V 0.0.3.5 implementing Sobel filters and unsharp masking filter	IE
QTFPX V 0.0.3.5 implementing featureextractlabeledarff	FE
QTFPX V 0.0.3.5 implementing classificationresultvisualizer	PA
VQI V 0.3.3	FE
Mindtct (part of NBIS V 4.1.0)	FE
Bozorth3 (part of NBIS V 4.1.0)	CL
sha256sum (GNU coreutils V 8.13)	CC

Table 76: Software inventory for the measurement support systems MS_1, MS_2 and the workstation WS_1

We omit the fileserver FS_1 from the Table 76 on the grounds that it is an infrastructure support only element and does not employ any sets of methods of the forensic process in digitised forensics SF. All of the sets methods of the forensic process for digitised forensics SF mentioned afterwards refer to the description in Section 4.3.2 of this thesis.

On the measurement support systems we install the software necessary to gather the data from the sensors S_1 (FRT CWL 600) and S_2 (Smiths Heimann LS1 LITE Xe), respectively. For the measurement support system MS_1 that amounts to the self-programmed wrapper DD+Acquire for the manufacturer supplied Acquire Software [FRT20, p. 1] for the data stream DS_{S1}. The software implements methods for the digitalisation of physical objects DPO and methods for the digitalisation of contextual information DCI. For the measurement support system MS_2 we use the acquisition software Smiths Heimann TestWizard for the LS1 LITE Xe livescanner [FCC+20] to gather the data and supply the data stream DS_{S2}. It represents methods for the digitalisation of physical objects DPO. With regards to the level of detail selection as demonstrated in Section 6 we have to conclude that the firmware of the Sensors S_1 and S_2 are a closely guarded secret containing the intellectual property of the respective manufacturers or their suppliers.

The workstation WS_1 is using a virtualised environment and is baseline-encrypted to ensure confidentiality of the highly person-related data. It acts as the examination environment after the data

gathering examination step DG_{SF} is finished. We use the image processing suite gimp [TGT20] for evaluation purposes. For the data investigation DI_{SF} we prepare our own QTFPX research tool suite implementing the Sobel filters and the unsharp masking filter as methods for image enhancement IE and the featureextractlabeledarff feature extractor as method for feature extraction FE (see [HKD+14, pp. 902808 - 902808-15]). We employ the Weka [UWa20] software suite as a provider of methods for classification CL. We further use classificationresultvisualizer from the QTFPX research tool suite to assemble a trace image based on the classification results as a method for the presentation and annotation of evidence PA.

For the data analysis DA_{SF} we install the NBIS software [NIS+20], which provides us with the methods mindtct for the feature extraction FE and with bozorth3 for the classification methods CL. We use those methods both to process the exemplary fingerprint gathered with the Sensor S_2 for evaluation purposes of our research. Further, we use them to provide us with evaluation means of the latent fingerprint processing to enable a matching against the exemplar fingerprint. However, this is only for scientific evaluation, we stand by our aim of not wanting to replace to forensic expert by providing fully automated analysis systems (see Section 3.4). Finally, we use the visual quality indices VQI [Hof20] for our research evaluation presented in [HKD+14, pp. 902808 - 902808-15] as a set of methods for feature extraction FE.

For the process accompanying documentation we use the sha256sum cryptographic checksumming program [LTP20] providing methods of chain of custody maintenance CC.

As depicted in Figure 63 by a dashed box, during the examination step of strategic preparation SP_{SF} we use forensic data types and sets of methods for the forensic process that are associated with case-specific examination steps. In the remainder of this section we provide a detailed description of methods that are exclusively not case-specific and can only take place during strategic preparation. We will address proceedings that can happen both during strategic preparation SP_{SF} and similarly in case-specific examination steps in their respective description.

We summarise step of strategic preparation SP_{SF} for digitised forensics SF as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DD of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DD of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 77:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
Weka V. 3.6.6 implementing SMO, J48, Bagging	DS_{S1}	DD_5, DD_6	MG	DD_7, DD_4	D_{IF}
GIMP V 2.6.12	DS_{S1}	DD_1	IE	DD_2, DD_4	D_{IF}

Table 77: Summary of the actions taken during strategic preparation SP_{SF} based on [HKD+14, pp. 902808 - 902808-15] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

The general idea of the approach presented in [HKD+14, p. 902808-2] is the division of an image representation of a latent fingerprint contained on a substrate into small blocks with the subsequent classification of each block belonging to either the substrate or the fingerprint. To train the

machine learning approach, we feed substrate characteristic feature data DD_6 and trace characteristic feature data DD_5 as data of the investigation target D_{IT} into the SMO, J48 and Bagging classifiers of the Weka software suite, which provides methods for model generation MG. We record the selection of classifiers best fitting our task as parameter data DD_4. Model data DD_7 form the data of the investigation result D_{IR}. To provide us with an approximation of a ground truth we use the differential image approach where an image of a substrate containing a latent fingerprint is subtracted from the same substrate without the fingerprint [HKD+14, p. 902808-5]. We feed the scans from the sensor as raw data DD_1 acting as data of the investigation target D_{IT} into the subtraction method provided by the gimp image-processing suite as a method for image enhancement IE. The resulting difference image is processed signal data DD_2 as data of investigation result D_{IR}. We record the settings for the gimp image-processing suite best fitting our task as parameter data DD_4 All of the aforementioned data types form part of the data containing forensically relevant information D_{IF} but no data containing case-specific, forensically relevant information D_{ISC}, since there is no specific case to examine, yet.

7.3.2 Operational preparation OP_{SF} for the use case of non-destructive latent fingerprint examination

During the examination step of operational preparation for digitised forensics OP_{SF} we prepare primary storage space to contain the data from the data gathering on the fileserver FS_1 and on the secondary storage space of the measurement support systems MS_1 and MS_2 (see Figure 64). Note, that the latter is only for evaluation purposes of our research. In a real crime case, this would occur if a suspect were available and asked to provide an exemplar fingerprint. Finally we initialise the storage space on the workstation WS_1. The sensor type selection is also part of the operational preparation OP_{SF}.

7.3.3 Data gathering DG_{SF} for the use case of non-destructive latent fingerprint examination

We summarise the step of data gathering DG_{SF} for digitised forensics SF as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DD of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DD of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 78:

Specific instantiation of the method of the forensic process	Accessed data streams DS	Forensic data type for data of the investigation target D_{IT}	Exploited method of the forensic process	Forensic data type for data of the investigation result D_{IR}	Involved sets regarding the information contained
DD+Acquire + FRT Acquire V1.46	DS_{S1}	DD_1	DPO	DD_1	D_{IF}, D_{IFC}
DD+Acquire + FRT Acquire V1.46	DS_{S1}	DD_1	DCI	DD_3	D_{IF}, D_{IFC}
Smiths Heimann TestWizard V1.10	DS_{S2}	DD_1	DPO	DD_2	D_{IFC}

Table 78: Summary of the actions taken during data gathering DG_{SF} based on [HKD+14, pp. 902808 - 902808-15] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained, dotted lines mark evaluation actions outside the forensic examination

During the examination step of data gathering DG_{SF} the digital data provided by the sensor S_1 (FRT CWL 600) in the data stream DS_{S1} representing raw data DD_1 as data of the investigation target D_{IT} is gathered by the measuring support system MS_1 using the our own wrapper application DD+Acquire and the manufacturer supplied FRT Acquire. In conjunction they provide methods for the digitalisation of physical objects DPO and return raw data DD_1 as the investigation result D_{IR}.

At the same time, our own wrapper application DD+Acquire and the manufacturer supplied FRT Acquire also accesses a different functionality of the sensor and gathers from the raw data DD_1 as data of the investigation target D_{IT} contextual data DD_3 about the measurement process as data of the investigation result D_{IR}. This functionality represents a method for the digitalisation of contextual information DCI. We disregard DD_3 for our experiments but document their presence during process accompanying documentation.

During our research we collect data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}.

The methods for DPO and DCI (dashed line in Figure 63) will also be executed as part of strategic preparation SP_{SF} to support the model generation (see Section 7.3.1). Here only data containing forensically relevant information D_{IF} and not data containing case-specific, forensically relevant information D_{IFC} are left since SP_{SF} is performed ahead of a specific incident.

For our evaluation purposes and as part of strategic preparation SP_{SF} we also gather exemplary live fingerprint data from the researcher that donates the fingerprint on the substrate (dotted section of the Table 78). We access data provided by the sensor S_2 in the data stream DS_{S2} representing raw data DD_1 as data of the investigation target D_{IT} using the Smiths Heimann TestWizard. It provides a method for the digitalisation of physical objects DPO and returns processed signal data DD_2 as the investigation result D_{IR} due to internal processing and removal of any background information. Thus, only data containing case relevant information D_{IFC} is gathered.

7.3.4 Data investigation DI_{SF} for the use case of non-destructive latent fingerprint examination

We summarise the step of data investigation DI_{SF} for digitised forensics SF as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DD of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DD of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 79.

During the step of data investigation DI_{SF} we preprocess (see also Section 3.4) the gathered raw data from our data stream DS_{S1} containing the substrate and latent fingerprint information. For this purpose we use our implementation of Sobel filters and the unsharp masking filters of our QTFPX tool suite (see [HKM+14, p. 902808-6]). Those filters represent methods for image enhancement IE and operate on raw data DD_1 as the data of the investigation target D_{IT}. The data of the investigation result D_{IR} are processed signal data DD_2 with sharpened contours. After the image enhancing, the data will still consist of both data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant data D_{IFC}.

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
QTFPX V 0.0.3.5 implementing Sobel and unsharp masking filters	DS_{S1}	DD_1	IE	DD_2	D_{IF}, D_{IFC}
QTFPX V 0.0.3.5 implementing featureextractlabeledarff	DS_{S1}	DD_2	FE	DD_5, DD_6	D_{IF}, D_{IFC}
Weka V. 3.6.6 implementing SMO, J48, Bagging	DS_{S1}	DD_5, DD_6	CL	DD_8	D_{IFC}
QTFPX V 0.0.3.5 implementing classificationresultvisualizer	DS_{S1}	DD_8	PA	DD_2	D_{IFC}

Table 79: Summary of the actions taken during data investigation DI_{SF} based on [HKD+14, pp. 902808 - 902808-15] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

Those IE methods (dashed line in Figure 63) will also be executed as part of strategic preparation SP_{SF} to support the model generation (see Section 7.3.1). Here only data containing forensically relevant information D_{IF} and not data containing case-specific, forensically relevant information D_{IFC} are left since SP_{SF} is performed ahead of a specific incident.

The resulting preprocessed signal data DD_2 from the data stream DS_{S1} act as data of the investigation target D_{IT} for the featureextractlabeledarff functionality in our QTFPX as a method for feature extraction FE. Here, as shown in [HKD+14, p. 902808-8] in total 408 features are calculated for each image block. Those features are both trace specific feature data DD_5 and general statisti-

cal and structural features that also represent substrate specific feature data DD_6 [HKD+14, p. 902808-7] and form the data of the investigation result D_{IR}. After this feature extraction, the data will still consist of both data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant data D_{IFC}.

Those FE methods (dashed line in Figure 63) will also be executed as part of strategic preparation SP_{SF} to support the model generation (see Section 7.3.1). Here only data containing forensically relevant information D_{IF} and not data containing case-specific, forensically relevant information D_{IFC} are left since SP_{SF} is performed ahead of a specific incident.

During feature extraction, in [HKD+14, p. 902808-7] we describe features that can be used to determine the presence of a fingerprint pattern. Those features could serve as a first qualitative indicator for *loss* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA, if they wrongly indicate the absence of a fingerprint pattern in a block as trace specific feature data DD_5 where our ground truth describes the presence of a fingerprint pattern (see also Figure 65 for the Venn diagram representation).

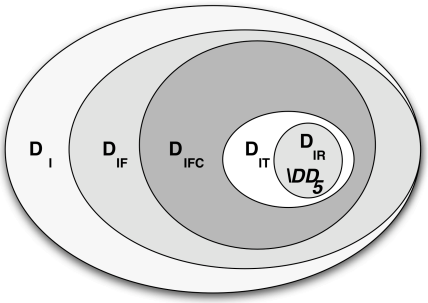

Figure 65: Loss due to the absence of DD_5 (denoted by the "\" symbol) as data of the investigation result D_{IR}, which directly conflicts with the ground truth from our experiments

Conversely, the presence of fingerprint patterns as trace specific feature data DD_5 in a block, where our ground truth shows its absence, could serve as a first qualitative indicator for *error* as postulated in Section 4.5 with regards to the Data-Centric Examination Approach DCEA (see also Figure 66 for the Venn diagram representation).

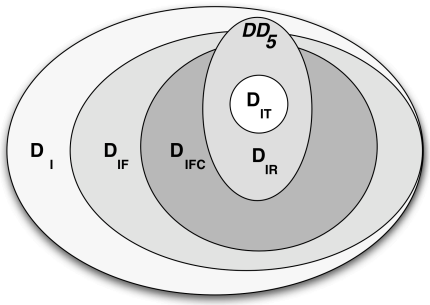

Figure 66: Error due to the presence of DD_5 as data of the investigation result D_{IR}, which directly conflicts with the ground truth from our experiments

The extracted trace specific feature data DD_5 and substrate specific feature data DD_6 from the data stream DS_{S1} form data of the investigation target D_{IT} for the methods for classification provided by the Weka suite, namely the SMO, J48 and Bagging classifier. Further, the model data DD_7 and the labels from the differential image approach based on DD_2 from strategic preparation SP_{SF} (see Section 7.3.1) are added. The data of the investigation result D_{IT} are classification result data DD_8 and ideally will only contain data containing case-specific, forensically relevant data D_{IFC}. Error and loss, however, as discussed above will influence the classification result.

Finally, using the classification result data DD_8 from the stream DS_{S1} as data of the investigation target D_{IT} the classificationresultvisualizer of our QTFPX is used as a method for the presentation

and annotation of evidence PA to assemble reconstructed a trace image as processed signal data DD_2 forming the data of the investigation result and ideally will only contain data containing case-specific, forensically relevant data D_{IFC}.

However, inherited *errors* from the previous steps in the classification result in an area of fingerprint pattern being present in the assembled trace image that are not there according to the ground truth (see [HKD+14, p. 902808-13]). Conversely, inherited *loss* from the previous steps in the classification result in an area of fingerprint pattern no being present in the assembled trace image that are there according to the ground truth (see [HKD+14, p. 902808-13]). This concludes our block based data investigation DI_{SF}. We will now move to the big picture and use biometric identification attempts during the following data analysis step.

7.3.5 Data analysis DA_{SF} for the use case of non-destructive latent fingerprint examination

We summarise the step of data analysis DA_{SH} for digitised forensics SF as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DD of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DD of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 80:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
Human examiner	DS_{S1}	DD_2,	CL	DD_8	D_{IFC}
Mindtct (part of NBIS V 4.1.0)	DS_{S1}, DS_{S2}	DD_2	FE	DD_5	D_{IFC}, D_{IF}*
Bozorth 3 (part of NBIS V 4.1.0)	DS_{S1}, DS_{S2}	DD_5	CL	DD_8	D_{IFC}, D_{IF}*
VQI V0.3.3	DS_{S1}	DD_1	FE	DD_5, DD_6	D_{IF}

Table 80: Summary of the actions taken during data analysis DA_{SF} based on [HKD+14, pp. 902808 - 902808-15] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained, dotted lines mark evaluations outside the forensic examination, dashed lines mark actions that are part of strategic preparation SP_{SF}, sets with * denote actions on latent fingerprints

The step of data analysis for digitised forensics DA_{SF} is, as stated in Section 3.4, performed by the human examiner. Here, the processed signal data DD_2 from the data stream DS_{S1} form the data of the investigation target D_{IT}. The human examiner employs intuition- and training-based approaches (see also Section 3.4.1), forming methods for classification. Tools to test examiner's bias have been suggested (see e.g. [Don20]). The data of the investigation result D_{IR} are classification result data DD_8, including the decision of match, non-match or inconclusiveness. In our research setup, where only case-specific fingerprints are deposited onto the substrates by the donors, the examiner operates on data containing case-specific, forensically relevant data D_{IFC}.

For our evaluation (dotted line in Figure 63 and Table 80) we also use a biometric matching provided by the NBIS biometric tool suite. We use our reconstructed fingerprint image as processed signal data DD_2 from the data stream DS_{S1} as data of the investigation target D_{IT} for the mindtct

functionality of the NBIS tool suite, which represents a method of feature extraction FE. As data of the investigation result D_{IR} we receive trace specific feature data DD_5. In this part we operate on data containing case-specific, forensically relevant data D_{IFC}, which due to errors and loss (see Section 7.3.4) could also included data containing forensically relevant information D_{IF}.

This is repeated with the as processed signal data DD_2 from the data stream DS_{S2} from the exemplary live fingerprint acquisition. We feed both trace specific feature data from DS_{S1} and DS_{S2} as data of the investigation target D_{IT} in the bozorth3 matcher as a method for classification CL. We receive as data of the investigation result D_{IR} classification result data DD_8. In this part we operate on data containing case-specific, forensically relevant data D_{IFC}.

As a further means of evaluation we already perform during strategic preparation SP_{SF} (see Section 7.3.1) as marked with a dashed line in Figure 63 and Table 80 we also evaluate the potential of visual quality indices VQI as a means to determine the sensor discriminatory power. We use the raw data DD_1 (one substrate containing no fingerprint and the same substrate with an added fingerprint) from the data stream DS_{S1} as data of the investigation target D_{IT} for the visual quality indices VQI as a method for feature extraction FE. Data of the investigation result D_{IR} returned from those methods are substrate specific feature data that DD_6 that denote the similarity of both data sources. Our evaluation does not include data containing case-specific, forensically relevant information D_{IFC} since SP_{SF} is performed ahead of a specific incident.

7.3.6 Documentation DO_{SF} for the use case of non-destructive latent fingerprint examination

We summarise the step of documentation DO_{SF} for digitised forensics SF as part of the sets of the examination steps applying the Data-Centric Examination Approach DCEA from Section 4 involving the forensic data types DD of the data of the investigation target D_{IT}, the accessed data streams DS, the exploited set of methods for the forensic process, the forensic data types DD of the data of the investigation result D_{IR} and the involved sets regarding the information contained from Section 4 in Table 81:

Specific instantiation of the method of the forensic process	**Accessed data streams DS**	**Forensic data type for data of the investigation target D_{IT}**	**Exploited method of the forensic process**	**Forensic data type for data of the investigation result D_{IR}**	**Involved sets regarding the information contained**
sha256	DS_{S1}, DS_{S2}	DD_1, DD_2, DD_3, DD_4 DD_5, DD_6, DD_7, DD_8	CC	DD_9	D_{IF}, D_{IFC}

Table 81: Summary of the actions taken during documentation DO_{SF} based on [HKD+14, pp. 902808 - 902808-15] with regards to the specific instantiation of the method of the forensic process, the accessed data stream DS, the forensic data type for data of the investigation target D_{IT}, the exploited method of the forensic process, the forensic data type for data of the investigation result D_{IR} and the involved sets regarding the information contained

As explained in Section 4.4.2.6, the examination step of documentation DO_{SF} is twofold. Across the whole examination process, including the strategic preparation SP_{SF}, all of the actions need to be recorded as process accompanying documentation for comprehensibility. That includes the settings used when digitising the physical object and the other configurable parameters in the signal processing and pattern recognition pipeline, which represent parameter data DD_4. Some measurement support systems include those parameter data in the digital object they create during acquisition. Every operation that manipulates a digital object with a file representation needs maintenance of its integrity. Cryptographic checksumming methods known to be secure against collisions at the time of their use and the foreseeable future can be used to show the maintenance of integrity. Each of the data types from all of the examination steps with a file representation

form the data of the investigation target D_{IT}. We use the sha256 algorithm as the method for chain of custody maintenance CC, which is deemed secure at the time of the research and the writing of [HKD+14, pp. 902808 - 902808-15]. The data forming the investigation result D_{IR} are chain of custody data DD_9. The sets involved in this method cover data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}.

The original article [HKD+14, pp. 902808 - 902808-15] and the report in this thesis with regards to the Data-Centric Examination Approach DCEA can be treated as different instances of the final report as part of the examination step of documentation DO_{SF}. They address a different audience with the first instance with a focus on technical details of the research and the second with a focus on forensic proceedings.

7.4 Summary of new findings and evaluation with regards to loss, error and uncertainty for digitised forensics in the use case of non-destructive latent fingerprint examination

In summary we can constitute that the examination of the use case can be described using our forensic data types DD for digitised forensics SF from Section 4.2.2, the sets of methods for the forensic process in digitised forensics SF from Section 4.3.2 and the sets of examination steps for digitised forensics SF from Section 4.4.1. We can use the separation of data of the investigation target D_{IT} and the data of the investigation result D_{IR} and the set relations between data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}. This description provides *common language* to describe different forensic examinations comparably and in a structured way, thereby allowing to systematically address issues of loss, error and uncertainty within procedures. The following Figure 67 summarises the examination.

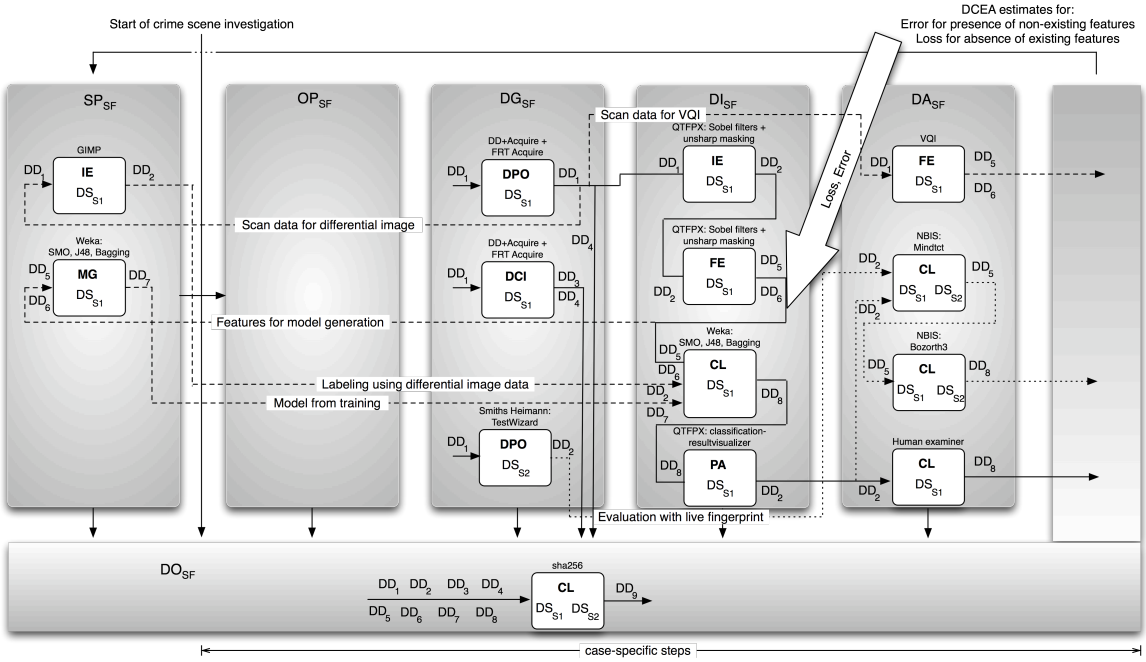

Figure 67: Depiction of the examination flow as summary of the use case of non-destructive latent fingerprint examination with the forensic data types and sets of methods for the forensic process

When describing the use case of non-destructive latent fingerprint examination using the Data-Centric Examination Approach DCEA, we can also use the data types for a first indicator for loss, error and uncertainty in a qualitative estimation as suggested in Section 4.5. As discussed in Section 7.3.4, the presence of blocks with features of fingerprint patterns and thus representing trace specific feature data DD_5 where there should not be any according to a ground truth, suggest the occurrence of *error* and thus could serve as a first qualitative indicator for it. The reason for such

error could be a highly structured substrate with patterns that share some characteristics of fingerprints and that contains the fingerprint pattern for our research.

The absence of blocks with features of fingerprint pattern representing trace specific feature data DD_5 where they should be according to a ground truth, pinpoint towards *loss* and thus could serve as a first qualitative indicator for it. The reason for such loss could be a faint fingerprint residue structure with respect to the contrast it offers against the substrate containing it.

It needs to be stated that our research work in [HKD+14] has the luxury of a ground truth derived from a differential image; otherwise this detection of those occurrences of error and loss could have gone unnoticed.

8. Comparison of the achievements for the chosen application examples and the applicability of the new Data-Centric Examination Approach for the forensic process

In this section we conclude our findings with regards to the Data-Centric Examination Approach DCEA introduced in this thesis and its application as a forensic process model allowing for a systematic, qualitative estimation of loss, error and uncertainty in digital and digitised forensic examinations. We start by summarising the requirements for the approach. Following that, we summarise our findings with regards to the formal description of loss, error and uncertainty and the contribution of a common language to describe forensic examinations by the Data-Centric Examination Approach DCEA. We reflect on the comparability of proceedings in an intra- and inter-examination context offered by the application of DCEA.

8.1 Application requirements

The application of the Data-Centric Examination Approach DCEA for digital and digitised forensics requires a data processing and the presence of *digital data* containing information both as target of the investigation and will return data containing information as a result. The approach will fail if the information is contained in analogous signal representations only (e.g. in the signal processing in the analogue domain prior to digitalisation). Loss, error and uncertainties arising from processes prior to digitalisation cannot be identified using the Data-Centric Examination Approach DCEA.

To be of assistance in providing qualitative estimates to loss, error and uncertainty in a case-specific examination in digital or digitised forensics apart from the trivial solution (full loss), the data forming the investigation target D_{IT} is required to include data containing case-specific, forensically relevant information D_{IFC} next to the expected data containing forensically relevant information D_{IF} and data containing information D_I.

The Data-Centric Examination Approach DCEA requires a system landscape analysis (Section 4.4.1.1) for an estimation regarding the forensic data types likely to be contained in the system under examination and thus being recoverable (at least in theory) and to determine system boundaries. It also requires the system landscape analysis of the examination environment to determine boundaries for the gathering, investigation, analysis and documentation of data.

Ahead of the application of the Data-Centric Examination Approach DCEA the context-sensitive definition of sets of examination steps, sets of methods for the forensic process and forensic data types according to the application area is required. The number and content of sets of examination steps and methods for the forensic process as well as the forensic data types can vary. In [AHK+19, p. 135] for example we propose to split the data type of user data into a new data type of application data and of functional data for the application field of industrial control systems ICS. Also the definition of substructures (e.g. subtypes of forensic data types) could be a solution for some application field.

The Data-Centric Examination Approach DCEA requires a level of detail selection as described in Section 4.4.1.2 and discussed in the use case of the examination of main memory in Section 5.4. This level of detail selections is crucial and sets the boundaries for the accuracy for the qualitative estimates towards loss, error and uncertainty for a given forensic examination. Much like the active intervention/alteration of data during the examination (see Section 3.3.1.3), this level of detail selection has to be documented and argued for as it represents a careful weighing process.

8.2 Summary of the findings

If the requirements from Section 8.1 are met, the Data-Centric Examination Approach DCEA can:

- provide *common language* describe different forensic examinations and their constituents in a structured manner,
- allow for an *intra- and inter-examination comparison* and depiction of the examination process,
- provide *qualitative estimates to loss, error and uncertainty* within procedures.

The sets of examination steps, sets of methods for the forensic process and the forensic data types provide a *common language* to describe the data of the investigation target D_{IT} and their processing using methods of the forensic process and resulting data of the investigation result D_{IR} and their composition ordered in time and space by the examination steps. They thus provide a hardware-/software-independent description and depiction of the examination. With those depictions, possible insertion points can be identified, e.g. if a method is retrospectively revealed to be erroneous or can cause loss or uncertainty or if the event reconstruction suggests looking for further inculpatory or exculpatory traces (see Section 2.7).

Provided that the instantiation for the forensic data types is the same, our approach allows for the *intra-examination comparison* for the methods used in the examination. This could be used as one decision criterion for the selection of a specific method (other criterions exist, see e.g. [HKD11, pp. 92-106]). Further, it could be used as a support for tool testing (see e.g. [Gray12] or [HML+11, pp. 1-6]).

If the instantiation of the examination steps and sets of methods for the forensic process match, an *inter-examination comparison* is also possible. Such a comparison employing the Data-Centric Examination Approach DCEA can be supportive in a comparative evaluation of the degree of maturity of the examinations or in questions regarding the evidentiary value.

The comparison of forensic data types expected to be returned by the execution of a method against a ground truth or common sense could be used as *qualitative estimates for loss, error or uncertainty*. This is even more descriptive if applied to all methods employed in a given examination. These findings could also be used to trigger a re-examination of an existing incident if new insights come to light, e.g. with a particular method being used etc. (see also Section 4.4.1.1 about the knowledge of manufacturers, examiners but also attackers).

In our experiments for all use cases in total we evaluated 31 methods from the sets of methods of the forensic process for both digital and digitised forensic. We covered all 6 sets of the examination steps and detected estimations for loss on 3 occasions, for error on 1 occasion and uncertainty on 2 occasions.

For the use case of the examination of main memory we were able to enhance the results of the forensic examination by reducing uncertainty with the proposition of using essential data from the registry. For use case of hidden data and device impersonation we identified sources of loss by missing the rearrangement of device storage capacity and subsequently could successfully detect instances of data hiding. We identified the problem of device impersonation regarding hardware information resulting in uncertainty. For the use case of non-destructive latent fingerprint examination we identified sources of loss and error based on the false classification of image blocks belonging to either the trace itself or the substrate.

Generally we detected the necessity of more advanced forensic methods instead of using methods never intended to be used in a forensic context and which have properties not forthcoming as tool usage in forensics (e.g. closed source, undocumented tools from questionable sources). Still, they returned better results as commonly accepted tools and thus highlight the need for further research in the identified gaps of method coverage in forensics.

Aside from the technical details we believe that the application of the Data-Centric Examination Approach DCEA can render forensic examinations comparable and could provide decision cri-

terions for forensic tool (read: method) selection, could provide arguments for the selection of a course of action and a successive evaluation of proceedings in the broadest possible application of forensic examinations ranging from technical support/troubleshooting to court trials.

9. Selected remaining future work

During our research into the Data-Centric Examination Approach DCEA we identified numerous sources for future work. They can largely separated into two directions; firstly the finer granularity of the description of the forensic process and thus adding more semantics and relations to the constituents of DCEA by formalising it further with a classification scheme for digital and digitised forensics. And secondly, we identify starting points to add a quantitative dimension. In the following we will elaborate on both directions of future work.

9.1 Classification scheme for examinations in digital and digitised forensics

In future we plan to establish a classification scheme based on tuples adding more information about forensic examinations enhancing comprehensibility at the cost of complexity. It is based on a list of tuples used for the formalisation to include accompanying factors. The tuple notation is motivated by by Matt Bishop (see e.g. [ZHR+07, p. 6]), a well-known and highly regarded researcher in IT security. The suggested classification scheme further supports a technology-independent classification as demanded by [Car06, p.145].

Similarly to the approach in Computer Forensic Tool Testing CFTT [Gray12], where forensic requirements categories and forensic requirements specification are defined, we propose for future work to add a number of accompanying factors for digital and digitised forensics. For the following description (Table 82 and Table 83) tuples are used to model the connection between sets of examination steps, sets of methods for the forensic process and forensic data types together with the additional accompanying factors, somewhat analogous to [ZHR+07]. This notation has the advantage of being easily extensible, should new requirements be identified.

$FE_{DF}= (SE_{DF}, CM_{DF}, DIT_{DF}, DIR_{DF}, HW/SW_{DF}, AN_{DF}, IL_{DF}, IP_{DF}, DR_{DF}, SI_{DF}, DV_{DF}, EV_{DF}, PM_{DF})$
SE_{DF} - Sets of examination steps (DCEA component as described in Section 4.4.1)
SM_{DF} - Sets of methods (DCEA component as described in Section 4.3.1)
DIT_{DF} - Data of the investigation target D_{IT} (DCEA forensic data types DT, Section 4.2.1)
DIR_{DF} - Data of the investigation result D_{IR} (DCEA forensic data types DT, Section 4.2.1)
HW/SW_{DF} - hard- and/or software (i.e. is the method available in software, hardware or a combination thereof)
AN_{DF} - Activation necessity (i.e. is the method readily available or does it require special actions on the system under examination)
IL_{DF} - Investigation location (refers to physical and logical locations, e.g. in networks)
IP_{DF} - Investigation precondition (measures that must or must not be taken ahead of the method)
DR_{DF} - data protection relevance (does the data of the investigation result D_{IR} contain personal data, see also the appendix Section 10.1.4)
SI_{DF} - structural impact (describes data and methods unavailable after the application of the recent method, Section 3.3.1.3)
DV_{DF} - data volume (corresponds to the storage capacity necessary to perform the method)
EV_{DF} - evidentiary value tendency (significance of the finding, (non-) essentialness of the data, see Section 2.7.2 and anti-forensics from Section 2.6 including deliberate placement of forged traces and attacks against the flow of the forensic examination to that end)
PM_{DF} - protective measures (alteration protection offered by the method for itself and investigation target and result against anti-forensics, Section 2.6)

Table 82: Tuple representation of a forensic examination in digitised forensics with first ideas for additional accompanying factors

A forensic examination in Digital Forensics (FE_{DF}) can be described as a 13-tuple containing the sets of examination steps, the sets of methods and the forensic data types and *additional accompanying factors* as first ideas in parentheses (Table 82), which need thorough evaluation.
For digitised forensics, a similar 11-tuple notation and corresponding *additional accompanying factors* for a forensic examination FE can be summarised as first ideas as shown in the following Table 83:

FE_{SF}= (SE_{SF}, IA_{SF}, DIT_{SF}, DIR_{SF}, HW/SW_{SF}, IL_{SF}, IP_{SF}, DR_{SF}, DV_{SF}, EV_{SF}, PM_{SF})
SE_{SF} - Sets of examination steps (DCEA component as described in Section 4.4.2)
SM_{SF} - Sets of methods (DCEA component as described in Section 4.3.2)
DIT_{SF} - Data of the investigation target D_{IT} (contains DCEA forensic data types DD, Section 4.2.2)
DIR_{SF} - Data of the investigation result D_{IR} (contains forensic data types DD, Section 4.2.2)
HW/SW_{SF} - hard- and/or software (i.e. is the method available in software, hardware or a combination thereof)
IL_{SF} - Investigation location (refers to physical and logical locations, e.g. in networks)
IP_{SF} - Investigation precondition (describes measures that must or must not be taken ahead of the method)
DR_{SF} - data protection relevance (does the data of the investigation result D_{IR} contain personal data, see also the appendix Section 10.1.4)
DV_{SF} - data volume (corresponds to the storage capacity necessary to perform the method)
EV_{SF} - evidentiary value tendency (significance of the finding, forgeability of digital media data, see anti-forensics from Section 2.6 including attacking the forensic processing flow and deliberate placement of forged traces [BFGK09])
PM_{DF} - protective measures (alteration protection offered by the method for itself and investigation target and result against anti-forensics, Section 2.6)

Table 83: Tuple representation of a forensic examination in digitised forensics with first ideas for additional accompanying factors

Although similar to the tuple for digital forensics, for digitised forensics we suggest to omit the "activation necessary, AN" item since the examination will only conducted on dedicated IT systems and thus do not need ad-hoc installations to support a forensic examination. Further, we can also omit the "structural impact, SI" item as a direct result of the digitalisation of crime scene traces and subsequent possible multiple runs of proceedings precisely due to the preceding digitalisation.

9.2 Starting points for a quantitative evaluation of examinations in digital and digitised forensics

We believe that the general idea of researching loss, error and uncertainty in digital and digitised forensics and providing a common language for intra- and inter-examination evaluation should be extended to contain quantitative elements.

First starting points into the research could include capacity/bandwidth determinations for all involved data streams, probably including techniques used in watermarking or steganography (see e.g. [Lan07] and [HiD15]). A second starting point could be the quantitative evaluations already performed in computer forensic tool testing. Here for the mass storage data stream the test specifications from Disk Imaging [NIST20] and Deleted File Recovery [NIST20a] could act as a starting point to derive quantitative measures. A further starting point especially for digitised forensics is provided by benchmarking (see [KHD+12]). For some branches of digitised forensics, a comprehensive categorisation for benchmarking containing lots of substructures, already exists

[KLD+11]. It too can serve as a source for quantitative measures to derive for sensor data streams.

Generally, the strategic preparation SP for both digital and digitised forensics, we believe, needs to be strengthened, as it deserves revaluation. One crucial aspect would be the need to train and certify examiners to aim for comparable (i.e. deterministic) examinations and highly standardised reports thereof. The core principles of forensic procedures including cross-evaluation components such as ACE-V should be extended (or at least designed extensible) and applied to other forensic procedures. In our opinion, the Data-Centric Examination Approach DCEA could be one such potential step towards this goal.

9.3 Further Research on the formalisation of loss, error and uncertainty

Sabah Jassim in the colloquium suggests the use of relations instead of transfer functions (see Section 4), as this would enable a more detailed description of the properties of the sets of data containing information D_I, data containing forensically relevant information D_{IF} and data containing case-specific, forensically relevant information D_{IFC}. Diverting from the definition of transfer functions as used in [Car03, p. 4] and, indeed, in this thesis, this could open up a promising path of research, especially in the estimation of loss, error and uncertainty but requires a complex restructuring of the modelling and the practical implementation.

9.4 Further research on the connection between configuration data DT_4 from digital forensics and parameter data DD_4 in digitised forensics and the general connection of digitised and digitised forensics

We thank Jana Dittmann for bringing up a very interesting topic in the colloquium regarding a possible connection between configuration data DT_4 from digital forensics DF and parameter data DD_4 in digitised forensics SF. They are very similar in the sense that they describe options that alter the behaviour of an IT system. We already propose (see Section 4.2.2) that the size of the configuration data DT_4 is larger than the size of parameter data DD_4 on the grounds that the set of configuration data of an IT System, e.g. when used as a measurement support system for digitised forensics (see Section 7.3.1), has more members as it contains data outside the parameterisation of the flow of examination steps (see Section 4.4.2.7) based on signal processing and pattern recognition. We see further future work in researching this connection and its implications.

Further connections between DT_4 and DD_4 are visible in the video surveillance scenario from Section 5, where the examination is at the crossroads between digital and digitised forensic examination and the examination could lead to either strictly digital forensics only proceeding, ignoring the semantic contents of the images themselves or the inclusion of the latter, involving digitised forensics and a description of proceedings, accordingly. We see interesting future work arising from this observation.

Another important aspect suggested by a connection between digital and digitised forensics and triggering future work is raised when an IT system used in digitised forensics SF is subject to an examination in digital forensics DF. Here we see future work e.g. whether the well-described examination process in digitised forensics can support the examination of the incident and, of course, we see future work based on the enormous implication of the potential compromise of IT systems used for digitised forensics. This importance in part originates from the latent nature of crime scene traces and their digital-only enhancement.

10. Appendix

The appendices contain parts of the thesis that are left out for the sake of brevity but are likely to further the understanding on the fringes of the Data-Centric Examination Approach DCEA introduced therein.

10.1 Additional legal and data protection requirements

Disclaimer: The whole thesis is not devised to provide a legal contribution. The author and the contributors are no legal experts. All facts described herein are only used to derive technical and procedural requirements for the contribution of this thesis

In this part of the appendix we provide additional information regarding the legal and data protection requirements that need to be adhered to during forensic examinations, including data protection and privacy issues.

10.1.1 Validation of evidence and the chain of evidence

The validity of forensic evidence directly affects its evidentiary value. A number of factors can affect the validity of the evidence used in digital forensics and thus add to loss, error and uncertainty (see Section 4.1), which are a very important part of the challenge addressed by this thesis. These include, according to [BHM08, p. 4], among others, the following:

- missing collection tools,
- failure to report exculpatory data,
- evidence taken out of context,
- misleading or false evidence,
- failure to identify relevant evidence,
- system and processing errors.

Thus, a separate step, namely the *validation of the evidence*, has to be executed. Using a step model (see also Section 3.1 and, indeed, Section 4.4), in [BHM08, pp. 4-5] the forensic process is split into two general domains, the investigative and the legal domain and the validation step is located in the investigative domain (Figure 68).

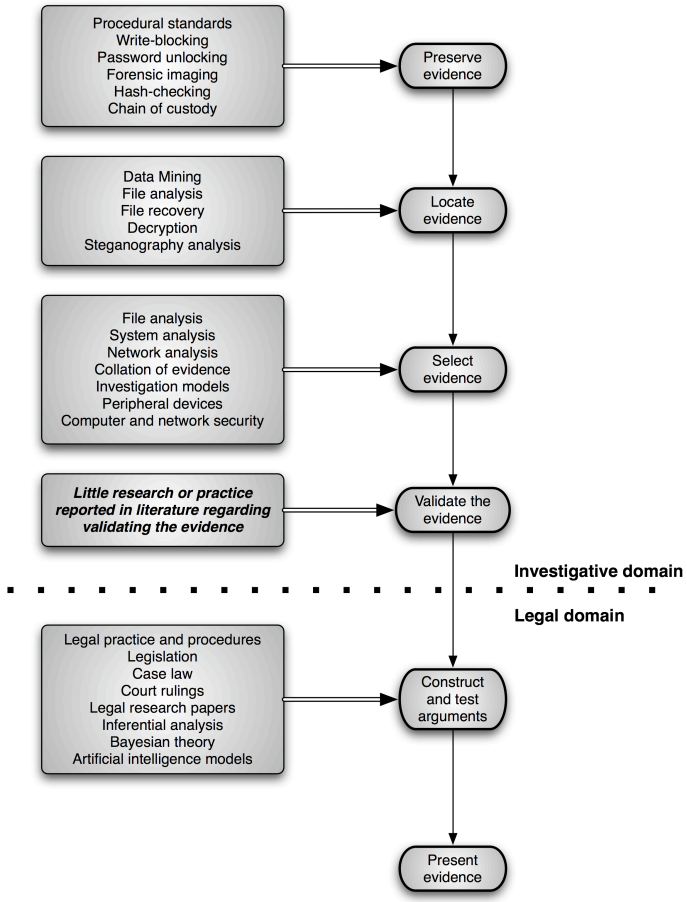

Figure 68: The need for evidence validation (derived from [BHM08, p. 5])

As pointed out by [BHM08, p.5], little research or practice is reported in literature regarding the validation of evidence. This thesis aims to find and evaluate qualitative measures that aid the validation process, especially in the light of loss, error and uncertainty.

To show how the validity of evidence can influence the reasoning and indeed, the outcome of a forensic investigation, one can look at the *chain of evidence*. Generally, in forensics the legal argument is based on logical chains of inferences, linking each piece of evidence to another with the strength of each inference used to determine the weight of the case [SiSh07, p. 180-181]. Using e.g. a timeline of reconstructed events (see also Section 2.2.2) or through the inferential process, typically the evidence together with its inferences is displayed in some form of graphic visualisation (see [SiSh07, p. 187-191]). Including the exculpatory evidence, a tree-like structure [BHM08, p. 9] emerges (see Figure 69).

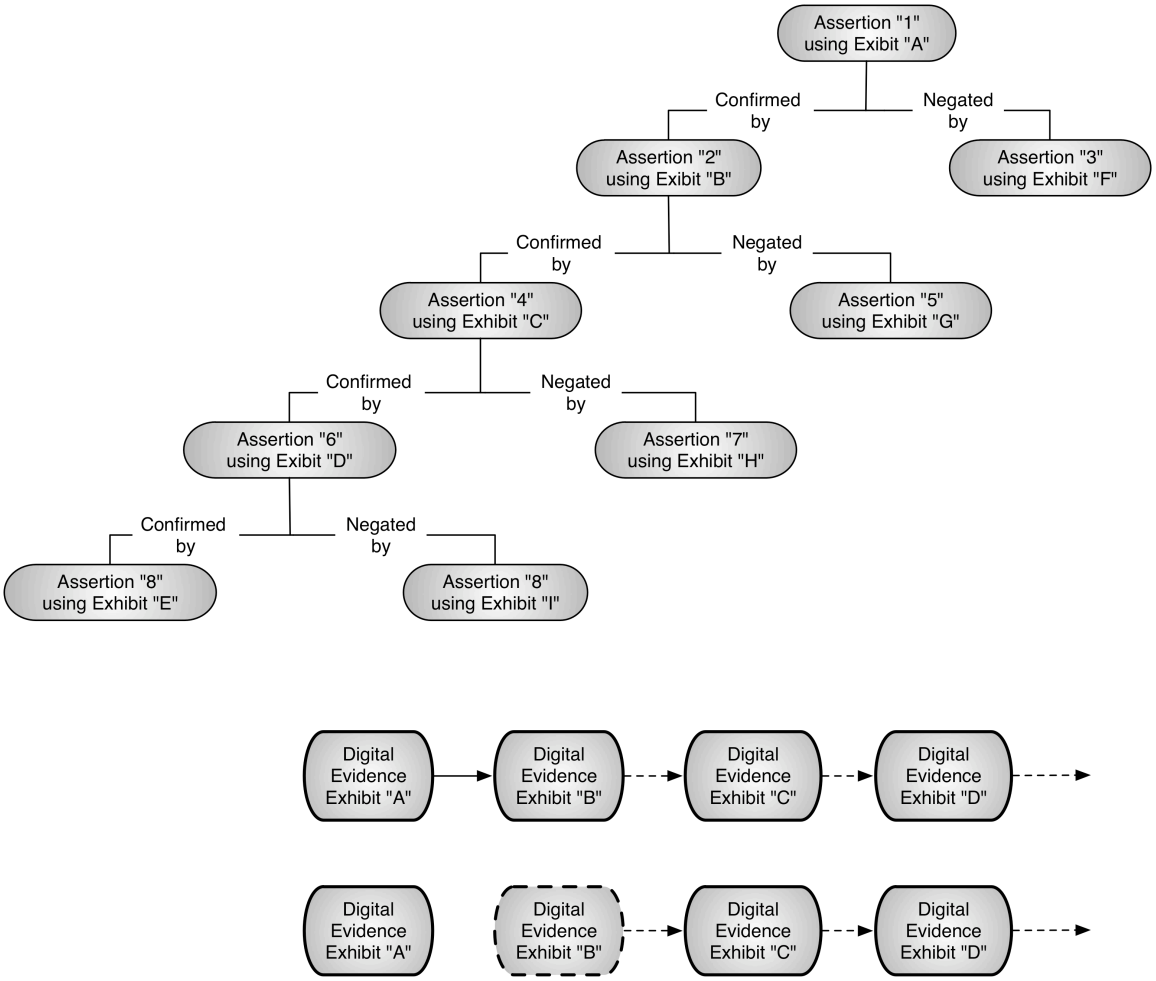

Figure 69: Chain of evidence based on the decomposition of the evidence (modified from [BHM08, p. 9])

After comprehensive and careful consideration all each horizontal nodes, a chain of evidence remains (middle part of Figure 69). However, if evidence is failing to pass the validation, the chain of evidence is broken (lower part of Figure 69).

The *validation process* for a particular digital evidence exhibit itself is depicted in the following Figure 70.

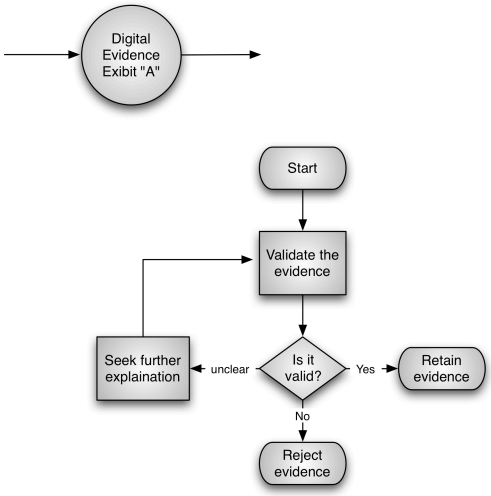

Figure 70: Validation process for one chain of evidence element (derived from [BHM08, p. 8])

This process is a very important tool to meet the demands placed on evidence to substantiate the Cyber Forensic Assurance [Dar10, pp. 61-64], in particular the component IIb of Authenticity/Original, which are outlined in Section 3.1.1. The Data-Centric Examination Approach DCEA introduced in this thesis in Section 4 provides a means to validate evidence, especially in the light of loss, error and uncertainty accompanying forensic investigations.

10.1.2 Types of forensic evidence

Forensic evidence used in a forensic examination can be grouped into three different types of evidence as circumstantial evidence, physical evidence and hearsay evidence, which are shortly outlined according to [New07, p. 8]:

- *Circumstantial evidence* is categorised as indirect evidence and is the result of combining seemingly unrelated facts that, when considered together, can be used to infer a conclusion. Such evidence is usually a theory supported by a significant quantity of corroborating evidence. Circumstantial evidence can include, partly, inferences about an event that was not seen. As stated in [BHM08, p. 7], digital evidence can be regarded as circumstantial evidence and is thus probabilistic in nature and, depending on the examination process, can be challenging and confounding observers in search for the truth regarding an issue.
- *Physical evidence* is the opposite of circumstantial evidence in that it cannot be wrong or perjure itself; only its interpretation can erroneous. Evidence of this type is direct, clear and tangible evidence of something, requiring no assumptions or added logic to prove that it is true. It is often collected from an eyewitness.
- *Hearsay evidence* consists of statements out of court from someone who is out not present to testify und oath and is based on what someone has told the witness and not on direct knowledge. Evidence of that type in general is not admissible, notable exceptions can be e.g. administrative-type hearings.

In digital and digitised forensics, data typically represents circumstantial evidence, since it has no physical presence and hard- and software needs to be used to render digital data perceivable.

10.1.3 Requirements for forensic evidence

In order for evidence to be admissible in a court of law, in a lot of countries this evidence has to meet certain requirements. In the following, the requirements for evidence in the United States of America are outlined since they are considered in number of other countries and because they underline the science part in the forensic sciences. Two important regulations, the Federal Rules of Evidences [Con12, pp.1-28] and the Daubert Hearing [DiG01, pp. 1-90] are outlined in the following.

10.1.3.1 The Federal Rules of Evidence (FRE)

In the USA the admissibility of evidence is governed by the Federal Rules of Evidence (FRE) [MoM11, p. 13-4]. Of particular relevance is Rule 702, Testimony by Experts [Con12, p. 14]:

"If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if

(1) the testimony is based upon sufficient facts or data,

(2) the testimony is the product of reliable principles and methods, and

(3) the witness has applied the principles and methods reliably to the facts of the case.

(As amended Apr. 17, 2000, eff. Dec. 1, 2000.)"

As stated in [MoM11, p. 13-6], the first statement implies that the expert's testimony rest on a sufficient basis that supports a reliable conclusion. Ideally, there would only be objective measurements and the facts or data would only yield unmistakeable answers. However, as pointed out in [MoM11, p. 13-6] it is the nature of science that some facts are not fully researched (thus leaving a grey area), which leads to some degree of subjectivity on the expert's opinion.

The second statement, according to [MoM11, p. 13-6], implies that it is not sufficient for the expert to explain his principles and methods. He also needs to show his knowledge about tests and research that demonstrates the reliability of the principles and methods.

The third statement, as pointed out in [MoM11, p. 13-6] implies that the expert has actually applied the knowledge of the second statements to the facts in the particular case.

Further, as stated in [HKD11, p. 94], especially for digital and digitised forensics the Federal Rules 901 and the Best Evidence Rules (Rules 1001-1008) are of relevance. In particular Rule 901, Requirement of Authentication or Identification, part b, clause 9 [Con12, pp. 22-23] states that "Evidence describing a process or system used to produce a result and showing that the process or system produces an accurate result", which would satisfy the requirement of authentication. As pointed out by [HKD11, p. 94], this requires an automated process. However, forensic examinations are not fully automated, which is where the "Article X. Contents of Writings, Recordings, and Photographs" [Con12, pp. 25-26] comes into play. This article consists of eight rules:

- Rule 1001, Definitions,
- Rule 1002, Requirement of Original,
- Rule 1003, Admissibility of Duplicates,
- Rule 1004, Admissibility of Other Evidence of Contents,
- Rule 1005, Public Records,
- Rule 1006, Summaries,
- Rule 1007, Testimony or Written Admission of Party,
- Rule 1008, Functions of Court and Jury,

which are also known as the Best Evidence Rules. Those rules define an original and regulate the admissibility of duplicates. This is important and in fact forms the basis of digital and digitised forensics. Rule 1001, clause 3 states [Con12, p. 25]:

"Original. - An 'original' of a writing or recording is the writing or recording itself or any counterpart intended to have the same effect by a person executing or issuing it. An 'original' of a photograph includes the negative or any print therefrom. If data are stored in a computer or similar device, any printout or other output readable by sight, shown to reflect the data accurately, is an 'original'."

This rule forms the basis of any forensic examination based on electronic evidence, since one important property of electronic evidence is, that although the data is stored with techniques from the physical world (e.g. magnetism), digital forensics looks at digital trace only it in its discrete representation an thus allows for perfect copies from this perspective.

10.1.3.2 Daubert Hearing and Daubert Factors

As a result in the court ruling in the case of Daubert vs. Merrell Dow Pharmaceuticals, it was stated that if a litigant challenges the admissibility of scientific evidence, it is the function of the trial court to act as a gatekeeper to determine if the evidence is relevant and reliable [MoM11, p. 13-13]. The *Daubert Challenge* [DiG01, pp. 1-90] is used since the Daubert decision by the U.S. Supreme Court for the admission of scientific expert evidence. This challenge evaluates three main criteria [DiG01, pp. 9-10]:

- Reliability: Is the evidence genuine, valid knowledge of the expert's field?
- Relevance: Will the evidence assist the trier of fact in determining a fact at issue?
- Qualifications: Does the expert have specialised knowledge in the field relevant to the testimony?

Also as a result of the Daubert decision, originally five *Daubert Factors* [DiG01, pp. 37-41] are used to assess, whether theories or methodologies are scientifically valid when evaluating evidence. Those five *Daubert Factors* assess the scientific validity according to:

- Peer review and publication,
- General acceptance in the relevant expert community,
- Potential for testing or actual testing,
- Known or potential rate of error,
- Existence and maintenance of standards controlling the use of the technique or method.

Especially the Daubert Factors "Potential for testing or actual testing" and "Known or potential rate of error" point towards the evaluation method of benchmarking, which is outlined in Section 10.6 of this thesis.

10.1.4 Data protection and privacy in forensic examinations

We stress the need for data protection in forensic examinations in Sections 4.1, 7.2 and include it in future work in Section 9.2. In the following we provide additional reflections of privacy and data protection considerations in forensic examinations.

Apart from the legal requirements about the comprehensiveness of a forensic examination and the validity of the theories and methodologies used, in many countries laws exist that govern the data protection and thus limit what an acting examiner is allowed to see and use.

Data protection regulations exist in a lot of countries. Although the detailed regulations differ from country to country, the underlying principles of data protection show a lot of similarities. We focus on the General Data Protection Regulation GDPR [EUC19, pp. 1-88] with a special attention paid towards the data protection in forensic examination.

Generally, since forensic examinations are all about digital evidence, some of that data (by its very nature) will be affected by the European Union General Data Protection Regulation (GDPR). The GDPR focuses exclusively on *personal data* with the goal of the protection of natural persons as a fundamental right and freedom.

In the context of forensic examinations, in [Jor19, p. 3] relevant definitions are identified and their applicability is discussed.

According to the definitions of the GDPR in Article 4 (1) [EUC19, p. 33]:

'‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;'

Personal data is processed, which according to the definitions of the GDPR in Article 4 (2) [EUC19, p. 33] is defined as:

'‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;'

The processing is determined by a controller, which according to the definitions of the GDPR in Article 4 (7) [EUC19, p. 33] is defined as:

'‘controller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;'

The processing is performed by a processor, which according to the definitions of the GDPR in Article 4 (8) [EUC19, p. 33] is defined as:

'‘processor’ means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;'

As stated in [Jor19, p. 9] all Digital Forensics and Incident Response (DFIR) will *process* potential *personal data* with most practitioners acting as processors. The GDPR defines *principles* in Article 5 [EUC19, pp. 35-36], out of which in [Jor19, p 11] are deemed relevant as personal data:

(a) 'processed lawfully, fairly and in a transparent manner in relation to the data subject (‘lawfulness, fairness and transparency’);

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (‘purpose limitation’);

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’);

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (‘accuracy’); (1) Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (OJ L 241, 17.9.2015, p. 1).

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject (‘storage limitation’);

(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures (‘integrity and confidentiality’).'

Of particular relevance is the *lawfulness of processing* of personal data [Jor19, p. 12]. The GDPR states in Article 6 that processing is lawful only if [EUC19, p. 36]:

(a) 'the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

(c) processing is necessary for compliance with a legal obligation to which the controller is subject;

(d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. '

[Jor19, p. 13] identifies the *conditions of consent* of the GDPR as relevant. Here the GDPR states in Article 7 [EUC19, p. 37]:

1. 'Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.
2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this regulation shall not be binding.
3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.
4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.'

According to [Jor19, p. 15], the relevant requirements towards the *processor* provided by Article 28 of the GDPR are [EUC19, p. 49]:

1. 'Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject.
2. The processor shall not engage another processor without prior specific or general written authorisation of the controller. In the case of general written authorisation, the processor shall inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes.
3. Processing by a processor shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that

sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller. In particular, that the processor:

(a) processes the personal data only on documented instructions from the controller, including with regard to transfers of personal data to a third country or an international organisation, unless required to do so by Union or Member State law to which the processor is subject; in such a case, the processor shall inform the controller of that legal requirement before processing, unless that law prohibits such information on important grounds of public interest;

(b) ensures that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality;

(c) takes all measures required pursuant to Article 32;

(d) respects the conditions referred to in paragraphs 2 and 4 for engaging another processor;

(e) taking into account the nature of the processing, assists the controller by appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the controller's obligation to respond to requests for exercising the data subject's rights laid down in Chapter III;

(f) assists the controller in ensuring compliance with the obligations pursuant to Articles 32 to 36 taking into account the nature of processing and the information available to the processor;

(g) at the choice of the controller, deletes or returns all the personal data to the controller after the end of the provision of services relating to processing, and deletes existing copies unless Union or Member State law requires storage of the personal data;

(h) makes available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller. '

According to [Jor19, p. 18], the relevant requirements *regarding records of processing activities* set by Article 30 of the GDPR are [EUC19, pp. 51-52]:

1. ' 1. Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information:

   (a) the name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer;

   (b) the purposes of the processing;

   (c) a description of the categories of data subjects and of the categories of personal data;

   (d) the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations;

   (e) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;

   (f) where possible, the envisaged time limits for erasure of the different categories of data;

   (g) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).

2. Each processor and, where applicable, the processor's representative shall maintain a record of all categories of processing activities carried out on behalf of a controller, containing:

   (a) the name and contact details of the processor or processors and of each controller on behalf of which the processor is acting, and, where applicable, of the controller's or the processor's representative, and the data protection officer;

   (b) the categories of processing carried out on behalf of each controller;

   (c) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;

   (d) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).'

As stated in [Jor19, p. 19], in order to comply with the above, practitioners need to ensure that they have appropriate mandates in place, together with suitable policies and procedures.
Further, [Jor19, p. 21] identifies the relevance of the need to ensure *security of processing* as demanded by Article 32 of the GDPR [EUC19, pp. 52-53]:

1. 'Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

   (a) the pseudonymisation and encryption of personal data;

   (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

   (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

   (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

2. In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.
3. Adherence to an approved code of conduct as referred to in Article 40 or an approved certification mechanism as referred to in Article 42 may be used as an element by which to demonstrate compliance with the requirements set out in paragraph 1 of this Article.
4. The controller and processor shall take steps to ensure that any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller, unless he or she is required to do so by Union or Member State law.'

The information collected and processed by the examiners, must be protected against access by unauthorised third parties. The main idea stated in [Sri07, p. 2] is that the necessity of forensic examination is acknowledged while at the same time the privacy rights of individuals must not be violated. This is even more relevant as it also includes the privacy rights of individuals not related to any suspected incident. Furthermore, following the suggested policies could lead to evidence acceptable in court. After this short legal detour in the following important basics of biometric systems are outlined, which in turn are relevant for digitised forensics.

10.2 Attacker Model

As stated in Section 2.6, anti-forensics plays an important role in forensic examination and should be part of the mental grammar of examiners. Putting aside the noble goal of detecting vulnerabilities and other shortcomings of methods of the forensic process and using the mindset of operating in an adversarial environment as outlined in Section 2.6 and in [Gar07], in [KHD11, pp. 95-96] we stress the need for a modelling of attackers, which includes the modelling of *attacker classes* based on [AnK97, pp. 125-126] and the *objective items* taken from the incident taxonomy [Hol12, pp. 1-32]. The attacker classes can be separated into three distinctive groups regarding their knowledge and equipment according to [AnK97, pp. 125-126]:

- Class I (clever outsiders): Members of this class are often very intelligent but may have insufficient knowledge of the system to be attacked. They may have access to only moderately sophisticated equipment and often try to exploit an existing weakness in the system, instead of trying to create one,
- Class II (knowledgeable insiders): Members of this class have substantial specialised technical education and experience and varying degrees of understanding of parts of the system but potential access to most of it. They often have access to highly sophisticated tools and instruments for analysis,
- Class III (funded organisations): Members of this class are capable of assembling teams of specialists with related and complementary skills. Those members can have great funding resources at their disposal. Knowledge-wise they are capable of in-depth analysis of the system, designing sophisticated attacks, and using the most advanced analysis tools. Also they can choose to include Class II adversaries as part of the attack team.

In general, in digital and digitised forensics, depending on the severity of the actions being forensically examined, it is prudent to assume that no class can be safely excluded.

Using the items contained in the *action* and *results* category (forming an event) and speculating by using the *attacker* and *objectives* category of the incident taxonomy from [HoL12, pp. 1-32] and reflecting on the anti-forensics outlined in Section 2.6 in the following, a few exemplary attacker models are outlined from [HKD11, pp. 92-106].

One such attacker model could include the alteration of data and the data-hiding targeting data by the action of modify. Potential attackers could be hackers, spies, professional criminals or even the examiner himself (in any of these roles). Objectives could be status or damage.

Another attacker model could include the stealing of data targeting data by the action of read. This becomes a valuable course of action if the examined system contains valuable data if the examiner has the objective of financial or political gain. Potential attackers could be hackers, spies, professional criminals or even the examiner himself (in any of these roles).

One last attacker model could address the forensic workstation by targeting the processes of a forensic workstation with the action of modify. The objective could be damage in order to cover up a previous attack. There might be various attackers: hackers, spies, professional criminals or corporate raiders.

10.3 Linking digital data to an individual and potential for forgeries

As discussed in Section 2.1.1, we abstain from a detailed discussion of multimedia forensics. However, the general discussion of the origin of the signals forming the origin of the raw data DT_1 for digital forensics and raw sensor data DD_1 and contextual data DD_3 for digital and digitised forensics, respectively and potential manipulation points throughout the subsequent processing flow of the examination (see also Section 2.6 about anti-forensics), we believe, deserve a mention.

A very important property of digital data and the information inferred from it is that this data does not contain any physical imprints, which connects the data to the individual who caused its gen-

eration or modification [Glad04, p. 16]. This means that on its own, there is no intrinsic link between digital data and a person. However, if the information contained in this data is acquired from the physical world, e.g. for biometric systems (see Section 2.8), this linkage can be established. Other means of tying data to individuals (see also [Vie06]) can be possession-based (e.g. a hardware token such as a smart-card) or knowledge-based (e.g. password as part of a digital signature). Such linkage of data to individuals is relevant to forensic examinations in two ways; electronic evidence needs to be tied to persons for event reconstruction (see Section 2.2) and the examiners need to ensure the authenticity and integrity of the electronic evidence.

In [BFGK09, p. 91] a further distinction is being drawn between digitised signal data used for multimedia forensics (and thus digitised forensics in the context of this thesis) and digital data created synthetically inside an IT system as part of the underlying finite automata principle (thus digital forensics in the context of this thesis) with regards to *forgeability* and *degrees of freedom*. The potential for forgeability (i.e. to craft an artificial piece of evidence with properties matching the demands of the counterfeiter) is high for digital data and lower for digitised forensics and at its lowest for conventional forensics. Contrary to that, the degree of freedom (i.e. the set of all possible objects with all possible combinations of properties) is at is lowest for digital data, higher for digitised data and at its highest for physical (analogous) objects. However, even for the digital data with its lowest degree of freedom, according to [BFKG09, p. 93], for a typical IT system in an abstract view, the storage medium of 100GB translates into a magnitude of $2^{10^{11}}$ potential states, which by far exceeds the number of atoms in the known universe. Thus, for digital data created synthetically (e.g. in a communication scenario), apart from high potential for forgeability, degrees of freedom exist.

10.4 Additional modelling of forensic data types for digital forensics for selected application domains

In this part of the appendix, we apply the data streams DS and corresponding data types for digital forensics (see Section 4.2.1) to selected application domains that are outside the exemplary use cases presented in Sections 5, 6 and 7 but underline the versatility of the data modelling already performed. We maintain the separation into the generalised structure and useful characteristics for forensics employed in Section 4.2.1.

10.4.1 Generalised structure of data organisation in mass storage devices for backup/archival purposes

A slightly different approach (using the same elements from the above) is used in tape drives and other sequential access methods. They are typically used for archival/long term storage purposes (i.e. backups). They are still in use (including the author of this thesis) due to a number of advantages (e.g. low cost media, dedicated use for backup purposes only, long data retention etc.). Thus, they also have a forensic relevance and evidence retrieved off this media is still used in court cases [Man19].

Tape drives (mostly due to their sequential access property) have a different partitioning. They do not employ a hierarchical file system structure. However, files can be grouped by partitioning the tape [Nik05, p. 5], also referred to as tape directories (see Figure 71). Please note, that the term files in the context of tape drives does not describe data entities contained in a user file system. Rather, the term denotes SCSI tape files as defined by the SCSI Stream Commands (SSC-3) standard (see e.g. [Sea19, pp. 225-226]). The user files are contained in the data entities specific to the backup application used (e.g. UNIX dump, Retrospect Backup, NTBackup, BRU etc.), which reside in instances of those SCSI tape files.

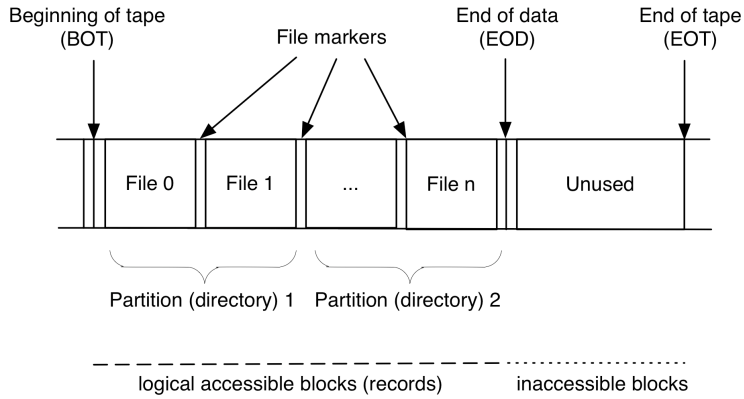

Figure 71: Simplified abstract view on tape media based on logical blocks, partitions and tape markers (summarised and extended from [Nik05])

The data area starts with a beginning of tape (BOT) marker and is bounded by an end of tape (EOT) marker. This is typically factory set and somewhat similar to the parent volume on other media (see also Figure 26, p. 67). However, the accessible area is limited further by the end of data (EOD) marker automatically set by the last recording process. Using the means provided by the tape drive, the tape cannot be spaced past this marker without opening a new writing session [Nik05, p. 15]. If the tape was re-used, data still present from an earlier backup session (potentially ranging from EOD to EOT) is not accessible although being still present. Depending on the particular tape technology, the directory containing the session log and the inventory of the tape are realised either on partition 1 of the tape or (in newer systems) on a separate chip contained in the tape housing [Nik05, p. 9].

Also, often for archival and backup processes, optical (re-) writable media are used frequently use in their implementation as CDs or DVDs and thus posses forensic relevance [Cro07, p. xxi]. They represent a mixture of random access during reads and sequential writes during the writing portion of media use. Generally, three modes of writing onto DVD/CD media are available [Cro07, p. 30-31]:

- Track-at-once - involves writing one track and shutting of the writing laser, resulting in two unreadable sectors on the disc, the table of contents (TOC) is constructed from track information and written automatically after finishing a session,
- Disk-at-once - involves creating the TOC first and writing the tracks without the unreadable sector gap,
- Incremental Recording or Packet Writing - sequential writes of small amounts of data without unreadable sector gaps but with added overhead of sectors for each packet, a packet corresponds either to the file to be written for CD-R or a fixed size of sectors for CD-RW/DVD-RW), a replacement of packets is possible.

Irrespective of the recording mode, the discs can contain purpose designed file systems such as ISO 9660 or Universal Disk Format (UDF) or file systems originally designed for random access media (e.g. the hierarchical file system, HFS, used in Apple Macintosh devices). A DVD is sector-oriented and has some special areas (see Figure 72) that form the final media.

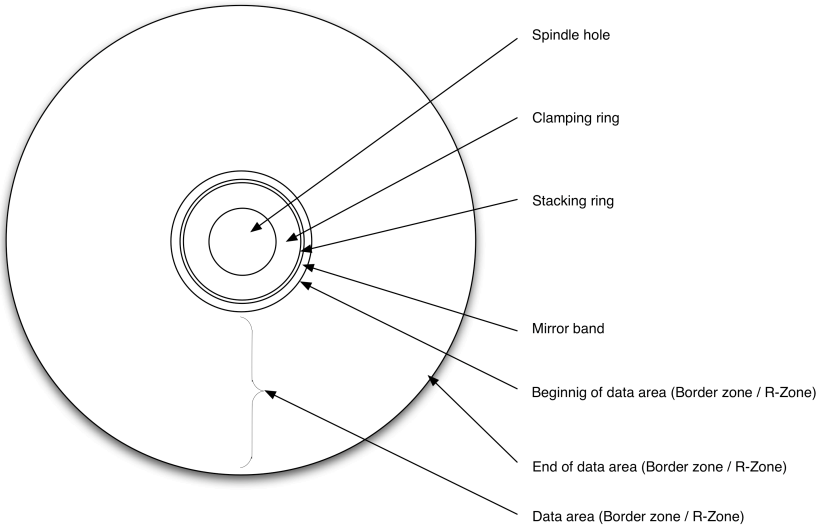

Figure 72: Abstract view of DVD structure (enriched from [Cro07, pp. 5-14])

These special areas carrying no data readable by the drive are [Cro07, p. 5]:

- Spindle hole (enables space-saving stacking of media),
- Clamping ring (definitely void of any information, preferred space for markings, including those of the investigator),
- Stacking ring (avoids adherence between the lacquer surface and the poly-carbonate of the following disc),
- Mirror band (contains content identifier, non-unique batch number or code).
- For the DVD (as opposed to the CD), the data is contained exclusively in the Border Zone / R-Zone.

A CD is also sector oriented, shares some of the special areas whilst being more complex in the data area (see Figure 73).

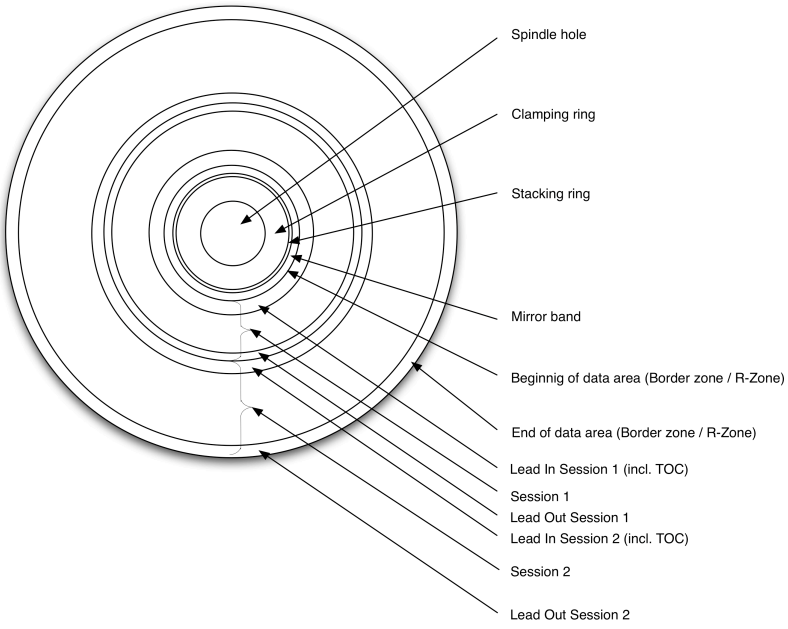

Figure 73: Abstract view on CD media structure (enriched from [Cro07, p. 5])

Additionally to the special areas carrying no data readable by the drive [Cro07, p. 5]:

- Spindle hole (enables space-saving stacking of media),
- Clamping ring (definitely void of any information, preferred space for markings, including those of the investigator),
- Stacking ring (avoids adherence between the lacquer surface and the poly-carbonate of the following disc),
- Mirror band (contains content identifier, non-unique batch number or code),

the data section is separated into sessions (data tracks), each containing [Cro07, p. 12-14]:

- Lead In (area serving as a container for the Table-of-Contents of the disc, in a multi-session disc implemented as a pointer towards the last Lead In of the currently last session),
- Content (data files, media files),
- Lead Out (indicates the end of a disc or session).

Multi-session CDs, which are not closed (which would disallow any additions but render the media readable in all CD-ROM drives), could contain data that are easily hidden (at least at a quick glance/triage, see Section 3.1.8) by adding another session to that disk.

10.4.2 Useful characteristics of mass storage management for forensics and mapping to forensic data types for backup/archival purposes

Backup systems based on tape media (see Section 4.2.1.1.1) offer interesting properties for forensic investigators. They typically come with their own *write blocker* (a switch that is physically sensed by the drive, preventing any accidental write operations, see e.g. [Nik05, p. 14]). Data contained on the tape can be accessed on a logical block level, which we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the Table 84 shows.
Tape devices can be *configured* (e.g. to employ hardware compression, sector size etc.), which we define as configuration data DT_4 according to the data type definition in Table 3 in Section 4.2.1, as the Table 84 shows.
Backups (often spanning multiple volumes) create *extra data* as meta data such as the media pool forming the backup set. We define them up as details about data DT_3 according to the data type definition in Table 3 in Section 4.2.1, as the Table 84 shows.
Backup systems based on tape media often record a *detailed log* of the whole backup process (including tape history, error counts, number of files and partitions etc., see [Nik05, p. 9]). Here session data can easily derived, both from the media itself or from a database maintained by the backup software operating the drive (see e.g. [Nik05, p. 7]). The log data is typically organised in backup session, which is why we define them up as session data DT_7 according to the data type definition in Table 3 in Section 4.2.1, as the Table 84 shows. Tape drives and media even form a special type of slack that is known as *tape file RAM slack* [Nik05, p. 12]. It entails the data found between the logical end of a tape file and the end of the last logical tape block. This area can be padded by both the operating system and the backup system containing arbitrary chunk of memory, which could yield some more data that we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1, as the Table 84 shows, which can be relevant to the forensic investigation.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Tape files can be accessed without any known logic determined by the backup program logic, this applies ever more to residual data (slack) padded by the backup solution
Details about data (DT_3)	Meta data describing other data	Backup set member information is kept as meta data typically by the backup solution
Configuration data (DT_4)	Modify the behaviour of the system and applications	Some features (e.g. hardware compression, encryption etc.) can be configured
Session data (DT_7)	Data collected by a system during a session	Backup sessions as logged by the backup solution itself and the tape backup media
User data (DT_8)	Contents created, edited or consumed by the user	The files to be backed up from to user's perspective

Table 84: Forensic data types as defined in [KHD09, pp. 2-3] derived from characteristics of tape-based archival systems

DVD/CD based archival/backup systems are by definition write protected in the case of CD-R/DVD-R media as the can only be recorded once but read many times (WORM characteristic). But even re-writable media such as CD-RW/DVD-RW need a dedicated write-capable device. By using a CD/DVD read-only device, the media is safe from accidental alteration.
DVD/CD based archival/backup systems also offer a multi session recording, the history of which can be retrieved using special forensic software (see e.g. [Cro07, p. 186]), delivering data that we define as session data DT_7 according to the data type definition in Table 3 in Section 4.2.1, as the Table 85 shows. Also, data hidden by adding a session, known as Multi-Session Hiding [Cro07, p. 167], can be revealed. Further, rewritable media, which have been quick erased (blanked) and thus, only the table of contents (TOC) area has been deleted, can be acquired using specially modified drives, specially recorded media (case/disc specific) and dedicated software. That way, the filesystem's raw data, which we define as raw data DT_1 and meta data concerning the filesystem within, which we define as details about data DT_3 can be gained. Both definitions are made according to the data type definition in Table 3 in Section 4.2.1, as the Table 85 shows.

Forensic data type	Description according to [KHD09]	Context in mass storage forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	With access to fast erased rewritable media, all unstructured content is accessible
Details about data (DT_3)	Meta data describing other data	With access to fast erased rewritable media, a full restoration of files and meta data is possible
Session data (DT_7)	Data collected by a system during a session	DVD/CD media can be written in incremental steps, each step represents a session

Table 85: Forensic data types as defined in [KHD09, pp. 2-3] derived from characteristics of DVD/CD-based archival systems

This concludes our visit to systems used for backup/archival purposes and we now look at further examples regarding selected network structures.

10.4.3 Generalised structure of data organisation in network forensics on the example of the CAN bus

The Controller Area Network (CAN) protocol used for intra car networking in automotive networks between nodes makes even more sparse use of OSI communication layers (see e.g. [Ric19, p. 1]). Here, electronic control units (ECU) as embedded devices (see [MöH19, p. 319]) communicate with each other as illustrated by Figure 74.

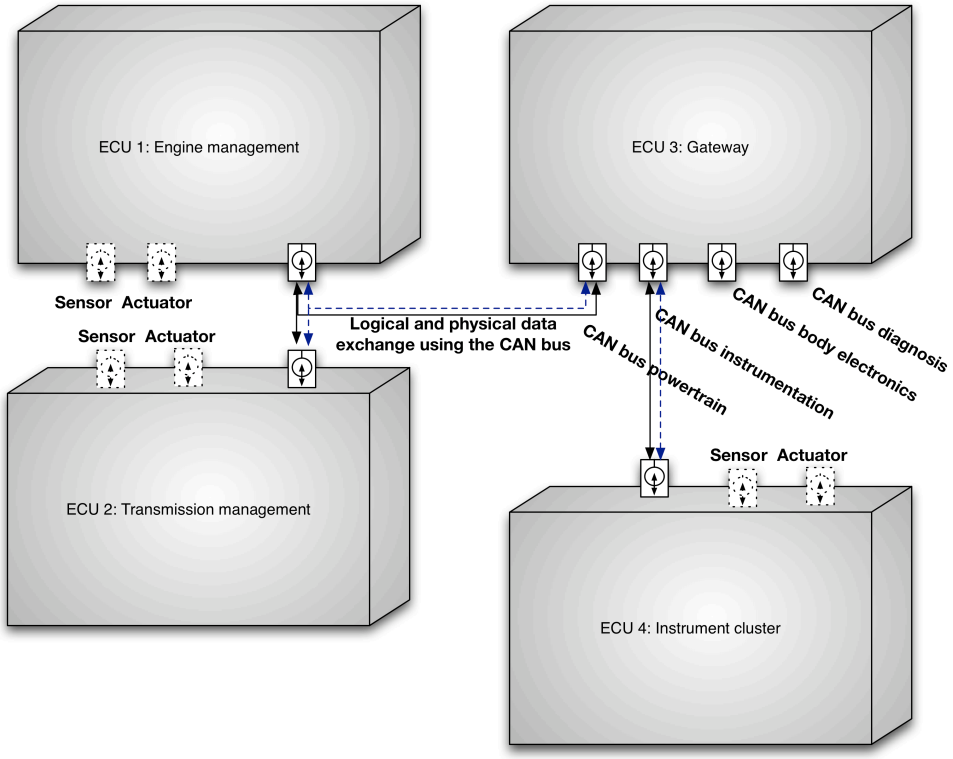

Figure 74: CAN bus communication between electronic control units (ECU) used for intra car networking

The CAN bus communication operates by daisy chaining of communication nodes that form a communication context in the shape of a sub bus (e.g. engine management ECU, transmission management as part of the powertrain sub bus). Those sub buses can be connected using gateways, which fulfil similar tasks as routers in conventional desktop IT networks. The CAN bus uses only the physical and the data-link layer of the OSI system (see [Ric19, p. 1]) and can be depicted as in Figure 75.

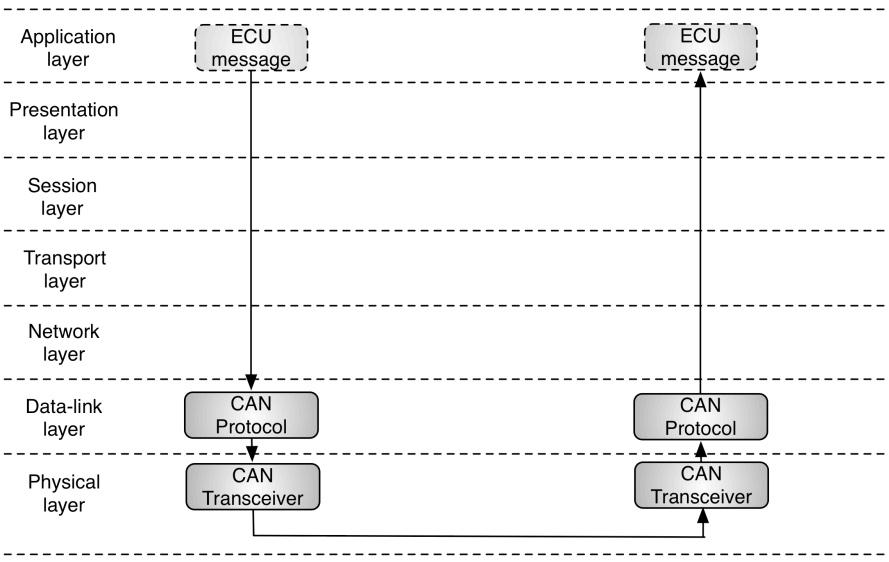

Figure 75: Generalised structure of data organisation in networks using the OSI/ISO reference model on the example of an ECU communicating to another ECU using the CAN bus protocol (inspired from [Cas11, p. 629] and using information from [Ric19, p. 1] and [MöH19, p. 114]) with the special status of the application layer (dashed box) taken from [Bos07, p. 78]

The layering of CAN communication is described in [Ric19, p. 1]) as follows:

- Physical layer: implements physical signalling (bit de/encoding, bit timing/ synchronisation),
- Data-link layer: establishes medium access control (MAC) by data (de-) capsulation and frame coding by (de-) stuffing as well as error coding/signalling and (de-) serialisation and establishes logical link control (LLC) by acceptance filtering and overload notification as well as recovery management.

In some publications (see e.g. [Bos07, p. 78]), the application layer is also added as:

- Application layer: consists of the application that processes and provides the information, only layer to be affected by user or sensor input.

We will use this structuring do derive forensic data types in the same manner as already shown in Section 4.2.1.3.2.

10.4.4 Useful characteristics of network management for forensics and mapping to forensic data types

As stated in Section 4.2.1.3.1, the CAN bus protocol only covers the physical layer, the data link layer and the application layer of the ISO/OSI reference model. Thus, using this protocol as source for structuring the forensic data types of the network data stream, we are limited by this rather small coverage of the ISO/OSI reference model. However, since access to the physical layer is possible, we can record data that we define as raw data DT_1 according to the data type definition in Table 3 in Section 4.2.1. Using the bus arbitration as a means for message prioritisation, we can define the CAN message ID as meta data and thus as details about data DT_3, according to the data type definition in Table 3 in Section 4.2.1. More than one CAN message with the same CAN message ID but with different payloads link the IDs into the same context, forming a session. We define those data contained in the CAN message as session data DT_7 according to the data type definition in Table 3 in Section 4.2.1. We define the payload of the CAN message as user data DT_8 according to the data type definition in Table 3 in Section 4.2.1, as it can transmit either directly visible data (e.g. street names in integrated navigation systems communicating results to the instrument cluster) or data that contains information for electronic control units (ECU) to act to the commands of the user (e.g. power window lifting mechanism). All the identified forensic data types are summarised in the following Table 86.

Forensic data type	Description according to [KHD09]	Context in network forensics characteristics
Raw data (DT_1)	A sequence of bits or data streams of system components not (yet) classified	Raw access to the bitstream on the CAN bus network wires
Details about data (DT_3)	Meta data describing other data	Bus arbitration and thus priority marker included to the CAN message
Session data (DT_7)	Data collected by a system during a session	Linkage of more than one CAN message payload to one another using the same CAN message ID
User data (DT_8)	Contents created, edited or consumed by the user	Application specific interpretation of the CAN message payload

Table 86: Forensic data types as defined in [KHD09, pp. 2-3] and their application in CAN bus networks

As we show with the above, our current modelling of the data types is also compatible with the CAN bus network. However, other networks may require other/additional data types.

10.5 Forensic evidence storage structures for digital and digitised forensics

As outlined in Section 2.1.1, digital and digitised forensics operates on electronic evidence, which needs to be stored secure whilst maintaining the chain of custody (Section 2.4). In Sections 4.4.1.6 and 4.4.2.6 respectively, we underline the necessity of process accompanying documentation for the examination as part of DO_{DF} and DO_{SF}. In the following, we provide some selected details about existing implementations that can greatly support and (partially) automate this process.

Several data structures, capable of storing the electronic evidence whilst maintaining the chain of custody, are devised. In the following, exemplary selected existing structures for digital and digitised forensics are outlined.

10.5.1 Forensic evidence storage structures for digital forensics

In the following two storage structures are outlined that, although sharing a number of similarities, are different in that one is intended for mainly for local use where the other supports distributed forensic examinations.

10.5.1.1 Digital Evidence Bag

Research into forensic storage structures includes Digital Evidence Bags (DEB). As outlined in [Bee09, pp. 24], these are designed to store provenance information related to the data collected. This is particular relevant in the cases of selective acquisition, where subsets of data from disparate sources, in which the source and the contextual data (i.e., the physical device and the subset of data that is not acquired) are no longer implicitly available and have to be explicitly retained. In addition, any explicitly retained data and information derived from it can be managed and thus can contribute information to the analytical process.

A Digital Evidence Bag as outlined in [Tur05, p. 225] is a wrapper for any type of digital based evidence or information. It provides (at least in theory) infinite capacity and can store data that is acquired both in live and post-mortem forensics (see also Section 2.1.1). Each bag contains its own tag information together with integrity information and continuity sections. As outlined in [Turn05, pp. 225-226] for each data acquisition three additional types of files (tag, bag, index) are created (see Figure 76).

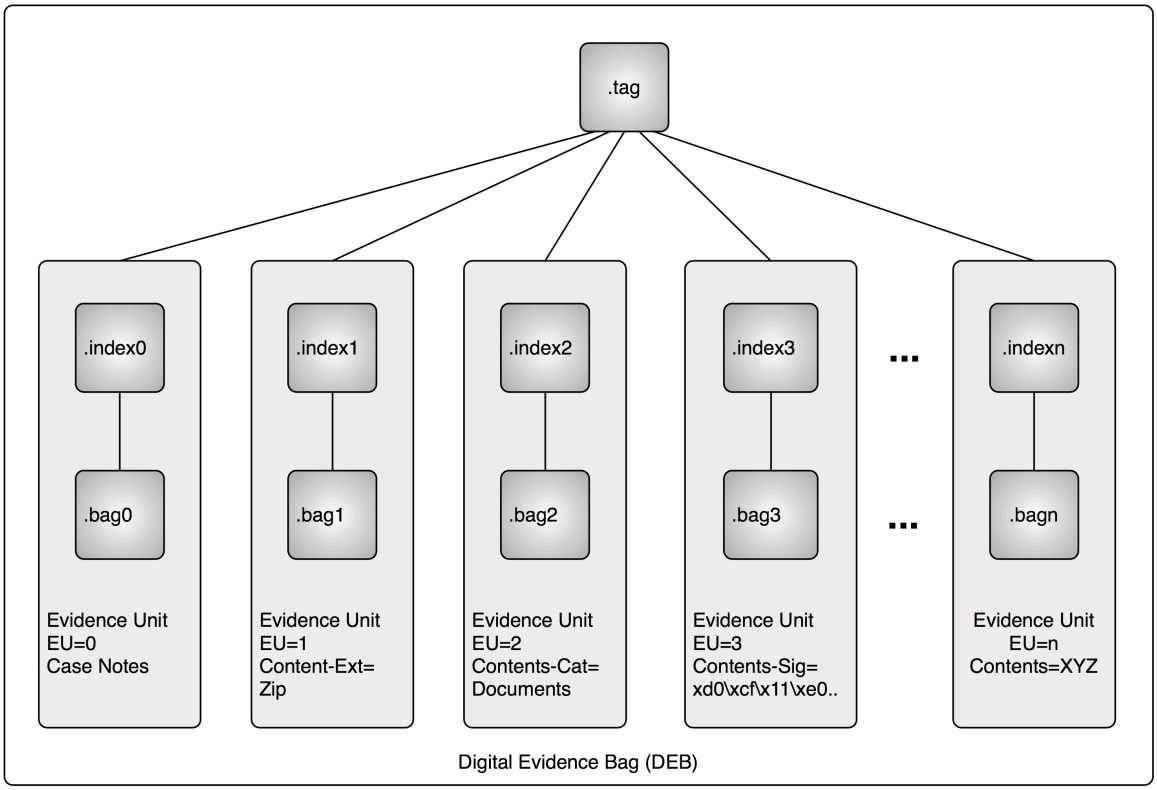

Figure 76: Digital Evidence Bag, modified from [Tur06, p. S62]

The root of the structure is formed by the *tag* file, which is a plain text file containing the information as outlined in [Turn05, pp. 225-226]:

- DEB reference identifier,
- details of the evidence contained in the DEB,
- name and organisation of person capturing the information,
- date and time the capture process started,
- a list of Evidence Units (EUs) contained in the DEB. An EU is the name given to an index file and its corresponding bag file,
- a hash of the captured information contained in the DEB,
- tag seal number comprised a hash of the tag file to date, this is equivalent to the traditional seal number,
- Tag Continuity Blocks (TCBs) containing continuity information of when any DEB application accesses the DEB,
- format definition of the .index file.

By updating the tag file with a Tag Continuity Block the DEB application reflects the history of operations performed on the bag file. It contains information such as date, the application a bag file is used with, an application signature to capture the category and version of the application. Also the tag seal number is updated by the DEB application. The *bag* file contains the acquired data, which can be raw binary data (e.g. dump data), files (e.g. as a result of logical copy), structured text (e.g. network packet recording), or categorised files (sorted e.g. according to their type). The *index* file corresponds to a bag file and contains metadata with details such as filenames, paths, and timestamps. But it can also contain data about the physical device that contains the evidence data such as type, serial number, and firmware version. Since the data contained in an index file is not fixed, the content type of those entries is specified in the format definition of the tag file.

10.5.1.2 Advanced Forensic Format 4

The Advanced Forensic Format 4 (AFF4) as introduced in [CGS09, pp. S57-S68] is designed to store multiple heterogeneous sources of data, which are likely to be expected in a forensic examination in digital forensics (see also Section 4.2.1). Such data can originate from different mass storage devices, network packets, main memory images but also extracted logical evidence and forensic workflow. Further, the file format is designed to be the basis of a global distributed evidence management system.

To accommodate for these capabilities, it employs an architecture, which is object oriented by providing a few generic objects with externally accessible behaviour. In doing so, the format encapsulates a number of functions, which are private and provides an interface to the forensic software that uses the AFF4 mechanism. This way, some requirements can be assured, regardless of the forensic application on top of the AFF4. The high level concept according to [CGS09, p. S59] includes the following items:

- An *AFF Objec*t is the basic building block of our file format. AFF Objects have a globally unique name as Uniform Resource Names (URN) as described in [SoM14], [Fie14] and [HMZ12]. The name is defined within the aff4 namespace, and is made unique by use of a unique identifier generated as per RFC4122 [LMS12].
- A *Relation* is a factual statement, which is used either to describe a relationship between two AFF Objects, or to describe some property of an object. The relation comprises of a tuple of (Subject, Attribute, Value). All metadata is reduced to this tuple notation.
- An *Evidence Volume* is a type of AFF Object, which is responsible for providing storage to AFF segments. Volumes must provide a mechanism for storing and retrieving segments by their URN. Currently, the Zip64 based volume and the Directory based volume are used.
- A *Stream* is an AFF Object, which provides the ability to seek and read random data. Stream objects implement abstracted storage, but must provide clients the stream like interface. For example, we discuss the Image stream used to store large images, the Map stream used to create transformations and the Encrypted stream used to provide encryption.
- A *Segment* is a single unit of data written to a volume. AFF4 segments have a segment name provided by their URN, a segment timestamp in GMT, and the segment contents. Segments are suitable for storing small quantities of data, and still present a stream interface.
- A *Reference* is a way of referencing objects by use of a Uniform Resource Identifier (URI). The URI can be another AFF Object URN or may be a more general Uniform Resource Locator (URL), such as for example a HTTP or FTP object. This innovation allows objects in one volume to refer to objects in different volumes, facilitating data fusion and cross-referencing.
- The *Resolver* is a central data store, which collects and resolves attributes for the different AFF Objects. The Resolver has universal visibility of objects from all volumes, and therefore guides implementations in resolving external references.

Most notably, next to the management of data originating from a multitude of sources and providing a basis for globally, distributed forensics, is the inclusion of methods to ensure the security aspect of *confidentiality* and in turn provide privacy protecting means (see appendix Section 10.1.4). The capabilities of the AFF4 are especially relevant for the approach introduced in Section 4 because of the support of multiple data sources whilst providing means to adhere to the demands of the chain of custody from Section 2.4.

10.5.2 Forensic evidence storage structures for digitised forensics

In the following two storage structures are outlined that, although sharing a number of similarities, are different in that the first is primarily designed to assist in one forensic examination in digitised forensics, namely digitised forensic dactyloscopy and the other is intended for replacing paper records, i.e. the paperless office.

10.5.2.1 Forensic container format for digitised forensic dactyloscopy

The forensic container for digitised forensic dactyloscopy as part of digitised forensics [KVL11, pp. 262-273] is a storage structure, which is intended to accompany a forensic examination and contain all the data created or modified during the examination. Furthermore, the requirements as laid out by the chain of custody and the documentation of the examination (Section 2.4 of this thesis) are addressed. The basis for the structure is the examination process in digitised forensic dactyloscopy. This process is derived from the biometric pipeline outlined in detail Section 2.8. In short, a biometric trait, in this case the fingerprint residue that forms a latent fingerprint, is digitised by the use of contact-less sensors, resulting in an *acquired* digital representation of aspects of this fingerprint residue. After a set of *pre-processing* steps, including image enhancement techniques using filters, *features are extracted*, which are then compared during the *classification* step. As introduced in [KVL11, p. 266], the *initialisation* of the forensic container can take place at the acquisition or at the pre-processing step, depending on whether the fingerprint data is acquired in the process or existing digitised fingerprint is used. Operations on the data during the pre-processing or the feature extraction steps are abstracted to *transformations* (see Figure 77).

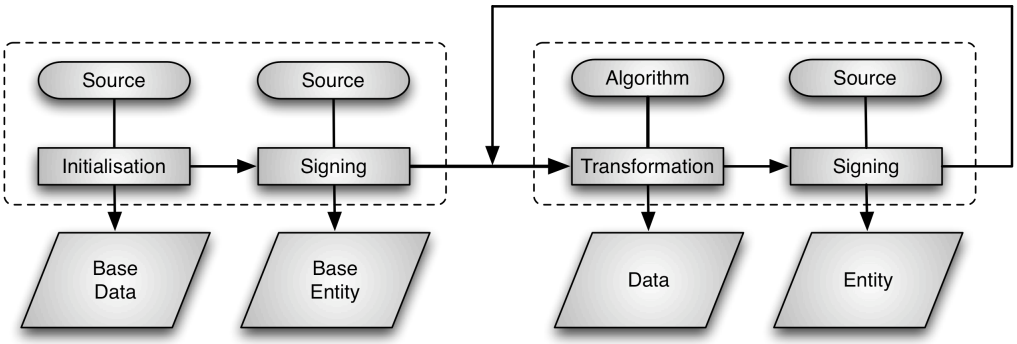

Figure 77: Generic initialisation and transformation process, modified from [KVL11, p. 266]

The resulting data of a transformation step is called an *entity* and every step depends on a *provenance*. The initialisation of the forensic container depends on a data source and a transformation depends on an *algorithm*. Transformations produce one entity as output und rely on at least one entity as input. The process of signing the data to form an entity supports the chain of custody by ensuring integrity and authenticity. The signing process is supposed to happen in a trusted environment (depicted by the dotted squares), whereas the resulting container can (in theory) be transferred over channels whose trustworthiness is unknown. Any unauthorised modification to items inside the container would result in failed security check. However, since fingerprint information represents highly person-related data, which remain unchanged over the lifetime of a person, it is also possible to encrypt the data, ensuring confidentiality and in turn also privacy (see also the appendix Section 10.1.4).

The meta model of the storage structure as introduced in [KVL11, pp. 266-267], which is modelled as a tree structure, defines elements and their relations (see Figure 78). The *container* element forms the root of this structure and is parent to a number of *editions*.

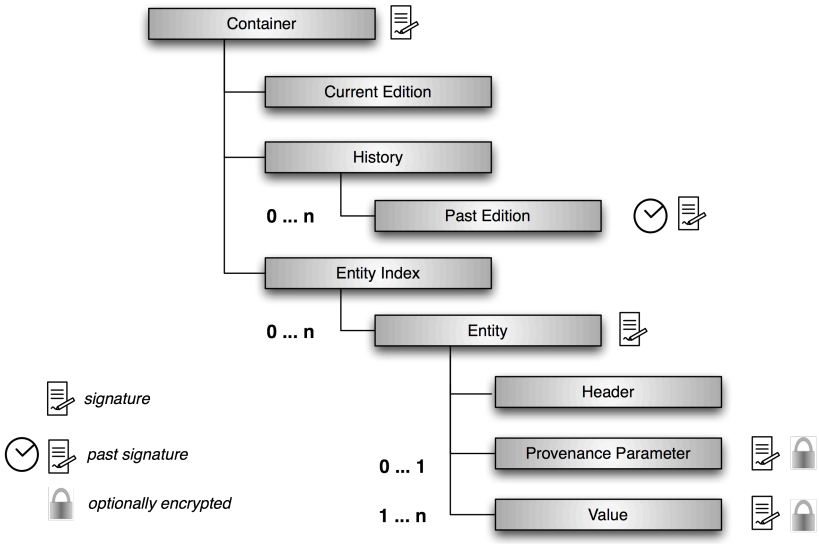

Figure 78: Meta model of the forensic container for digitised forensic dactyloscopy, modified from [KVL11, p. 268]

Keeping not only the current edition but also a stack of *past editions* embodies the chain of custody functionality. The stack of past editions forms the *history*. Obviously, when the container is extended, the former current edition is wrapped into a history item. An edition consists of a global unique identifier, a timestamp, the description of an *owner* creating or extending the container and a list of all entities added to the container during its creation or extension. The entity itself is parent to the *header* (a container-wide unique ID, a list of the IDs of all predecessor entities and a reference to the provenance of the entity), the *provenance parameter* (settings applied for a transformation) and at least one *value*. Further, the header references an entity type, whose definition describes all mandatory and optional values of the entity and the type of data. Similar to the AFF4 structure outlined in Section 10.5.1.2, the storage facility itself can be a folder structure or a ZIP archive. Entity values and parameter sets are stored as binary streams forming a file per value or parameter set whereas the other elements of meta model are stored as XML files.

10.5.2.2 Forensic Information Data Exchange Format (FIDEX)

The Forensic Information Data Exchange Format (FIDEX) is a modular, portable, NIEM-conformant, reusable XML data format with supporting documentation for use by criminal justice agencies wishing to share forensic information and data electronically [For12, p. 1]. It is based on the National Information Exchange Model (NIEM), whose key concepts are the *data components*, the *NIEM core*, the *domains*, the *communities of interest* and the *Information Exchange Package Documentation* [NPMO14, pp. 4-13].

In NIEM, according to [NPMO14, p. 5], the fundamental building block is the *data component*, which represents real-world objects and concepts (e.g. people, places, material things, events). Frequently and uniformly used components are specified in NIEM for reuse by practitioners irrespective of the operational exchange context, provided that they are semantically consistent. Data models, databases, data dictionaries, schemas and exchanges are examples of sources for data components, which are represented as of now by XML Schema. This way, a consistent definition and transmission of information exchange packets (IEPs) is possible. Components form a cohesive data model, which provides consistent semantics and structure.

Universal components, according to [NPMO14, pp. 6-7], are data components that are universally shared and understood, by (close to) all domains (e.g. person, address, organisation), i.e. carry the same meaning across all communities of interest. They form the *NIEM Core* and this set of universal components is stable and relatively small.

A *domain*, according to [NPMO14, pp. 7-9], refers to a business enterprise that broadly reflects the agencies, units of government, operational functions, services and information systems that are organised or affiliated to meet common objectives. Each domain traditionally includes a cohesive group of data stewards who:

- are subject-matter-experts (SMEs),
- have some level of authority within the represented domain,
- participate in harmonising conflicts and resolve data-component ambiguities.

Communities of Interest (COIs), according to [NPMO14, pp. 9-10], must have a shared vocabulary for the information to be exchanged. They are collaborating groups and exchange information. Such information exchange is documented by the reuse of data components and artefacts found in NIEM. COIs can coordinate the development of new domain content.

Information Exchange Package Documentations (IEPDs), according to [NPMO14, p. 12], include XML schemas that use or correctly extend NIEM components and define a class of XML exchange instances, subset schema want lists and style sheets. Further, they include documentation of the implementation of an information exchange packet using the schemas and other documentation (e.g. business requirements, domain models, use-case models). Also they include IEPD artefacts, which contain a manifest (list of artefacts in the IEPD) and the meta data registered with the IEPD used for e.g. indexing, search, discovery, maintenance, registration.

As outlined in the beginning, the FIDEX data format consists basically of two Information Exchange Package Documentations (IEPDs) for *forensic case submission* and *forensic case disposition* and include the XML schema and documentation on how to implement the IEPDs [For12, pp. 2-3]. As stated above, those FIDEX IEPDs are intended to replace paper-based forms in *forensic case submission* and allow an agency to transmit details about an incident together with evidence and examinations requested on that evidence. The IEPD for forensic case disposition is designed for the exchange between a crime laboratory and either a prosecutor management system or a court case management system. A change in the disposition status is received by the crime laboratory and can be used to determine whether outstanding examinations can be cancelled and removed from the backlog of cases.

After outlining a selection of exemplary chosen forensic evidence structures and their utility in maintaining a chain of custody for the evidence from forensic examinations in digital and digitised forensics, in the following an important tool to provide a common language for incidents and for forensic examinations is outlined, much like providing a common language for the exchange of evidence, which those forensic evidence structures provide.

10.6 Benchmarking of forensic tools and methods

In Sections 4.4.1.1 and 4.4.2.1 we state that tool testing and benchmarking of forensic tools and methods is an important part of strategic preparation for both digital forensics SP_{DF} and for digitised forensics SP_{SF}. In the following we provide additional information sectioned accordingly.

As outlined in Section 10.1.3.2 of this thesis, the Daubert factors [DiG01, pp. 37-41] play an important role in the decision of the admissibility of evidence in court. Three of the originally five Daubert factors point towards a benchmarking of forensic tools and methods:

- Potential for testing or actual testing,
- Known or potential rate of error,
- Existence and maintenance of standards controlling the use of the technique or method.

As for the last mentioned Daubert factor, the relevance towards benchmarking results from the fact that standards controlling the use of a technique or method asks for testing what the limits of that technique or method are and how and in which limits it is beneficial to use that technique or method.

Benchmarking, according to [SEH03, p. 76], refers to a test or set of tests used to compare the performance of alternative tools or techniques. Benchmarking (sometimes also called tool testing) is already used in digital and digitised forensics for the above named reasons and to improve existing solutions, which is further discussed in Section 4.4 of this thesis.

In the following, for both digital and digitised forensics an exemplary chosen benchmarking scheme is outlined.

10.6.1 General Test Methodology for Computer Forensic Tools in digital forensics

The National Institute for Standards and Technology (NIST) of the United States of America runs the Computer Forensic Tool Testing projects in order to provide a measure of assurance for the software tools used by law enforcement in computer forensic examinations [Gray12, p. 1]. The test methodology outlined in the following is based on the ISO/IEC 17025 "General Requirements for the Competence of Testing and Calibration Laboratories" [ISO99]. It entails the following tasks [Gray12, p. 1]:

- establish categories of forensic requirements,
- identify requirements for a specific category,
- develop test assertions based on requirements,
- develop test code for assertions,
- identify relevant test cases,
- develop testing procedures and method,
- report test results.

As outlined in [ISO99] and in accordance with the Daubert factor of 'Peer review and publication' [Gray12, p. 2] states that for cases where there is no standard test method, the results of each step is be made available for public review so that this process is an open, public process which incorporates and reflects the needs of a wide variety of law enforcement practitioners and suppliers of computer forensic tools.

Forensic requirement categories (1) represent groupings of forensic functions that are determined by expert users [Gray12, p. 2]. This grouping provides a smaller set of requirements that can be systematically approached by specialised forensic experts for testing whilst a narrowing of the scope allows for the identification of requirements for each functional grouping.

Forensic requirements specifications (2), according to [Gray12, p. 2], prescribe the technical and functional requirements to be fulfilled by a product. It identifies the requirements and features applicable to a category of forensic tools that are to be tested. An initial identification of a list of requirements or specifications from the category of forensic functions is made by a group of experts from federal, state and local aw enforcement organisations. The final requirements, however, are based on a consensus review from the communities using the tools. These requirements serve as a basis from which test assertions and test cases are developed that are used for forensic tool testing.

Test assertions (3) are statements of behaviour, action or condition that can be tested or measured [Gray12, p. 2]. In doing so, they connect the narrative of the specification with the test cases. A test assertion is an independent, complete, testable statement for a requirement in the specification and results in the realisation of one or more test cases. Interestingly, [Gray12] makes no assertions about the *development of test code* (4). However, such code is already published [NIST20].

Test cases (5), according to [Gray12, p. 2], specify what is to be tested or one instance of what is to be tested. The number of test cases that can be selected is limited by economic considerations. To narrow down the number of test cases, public review and the opinions of the experts from both computer science and forensic practitioners are used. Furthermore, the development of an independent experimental design by the NIST Statistical Engineering Division assists in the selection of test cases.

The *test method* (6) is a combination of the software used for testing and the procedures for completing the testing [Gray12, p. 2]. In other words, it describes how the testing is to be accomplished. The test method is the key to providing reasonable and practical assurance, that the forensic tools will meet the requirements of the investigators. This is outlined by the ISO/IEC 13210 "Information Technology – Test Methods for Measuring Conformance to POSIX" [IEEE98], which defines test methods as "the software, procedures, or other means specified by a POSIX standard to measure conformance. ... Test methods are intended to provide a reasonable, practical assurance that the implementation conforms to the standard." As stated in [Gray12, p. 3], the ISO/IEC 17025 [ISO99] specifically asks for documentation in the 'non-standard test method' case. To develop the requirements for the forensic tool testing method documentation, 16 generic items are used [Gray12, p. 3]:

- appropriate identification -- unique identifier which identifies the precise test case,
- scope -- the test method for which category of forensic requirements,
- description -- what product is being tested including version numbers,
- parameters -- the variables that are used to define the test, such as the size of the disks being tested, access, action, etc.,
- apparatus -- the testing environment or computers being used,
- reference standards -- the testing software or support software used,
- environmental conditions -- where the testing was completed,
- description of the procedure --
- identifying and documenting which equipment is being used and any preparation of the equipment before testing,
- checks to be made before testing begins including setup procedures,
- identifying how the documentation will be kept, name of files, etc.,
- identifying any procedures needed to protect the integrity of the test results,
- criteria for approval – what should be the expected results,
- data – what data will be captured, how the results will be analyzed, and presented.

If the test method is not valid, there can be no assurance of integrity, reliability and correctness of test results as stated by international guidelines as outlined by Section 5.4.5 of ISO/IEC 17025, "Validation of methods" [ISO99]:

"Section 5.4.5.1 Validation is the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled

Section 5.4.5.2 The laboratory shall validate non-standard methods, laboratory-designed/developed methods, ... to confirm that the methods are fit for the intended use. The validation shall be as extensive as is necessary to meet the needs of the given application or field of application. The laboratory shall record the results obtained, the procedure used for the validation, and a statement as to whether the method is fit for the intended use. ..."

The approach in [Gray12, p. 4] is that all non-standardised procedures used and all testing software will be made available for any interested party to use and judge.

Test result reporting (7) is all about the presentation of the test results [Gray12, p. 4]. As stated in [ISO99]:

"The results of each test, calibration, or series of tests or calibrations carried out by the laboratory shall be reported accurately, clearly, unambiguously and objectively, and in accordance with any specific instructions in the test or calibration methods. The results shall be reported, usually in a test report ..., and shall include all the information requested by the client and necessary for the interpretation of the test or calibration results and all information required by the method used ... "

According to [Gray12, p. 4] the forensic tool test results should include:

- title stating what product was tested,
- identification of the testing environment, i.e., where the tests were run,
- unique identifier for the test report, that identifier will be repeated on each page in order to ensure that the page is recognised as a part of the test report,
- the name and address of the vendor,
- identification of the testing software used,
- unambiguous identification of the product tested including version, patches, etc.,
- the test with the criteria for measurement,
- the name(s), function(s) and signature(s) or equivalent identification of person(s) authorising the test report,
- where appropriate and needed, opinions and interpretations,
- additional information, which may be required by specific methods, clients, or groups of clients.

An important fact of any benchmarking is the *repeatability and reproducibility* of the benchmarking results. The procedures of the test method in addition to the testing software ensure the repeatability and reproducibility [Gray12, p. 4]. ISO 5725 "Accuracy (trueness and precision) of measurement methods and results" [ISO94] and [ISO94a] define those very important terms as follows:

- *repeatability*: Precision under repeatability conditions.
- *repeatability conditions*: Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.
- *reproducibility*: Precision under reproducibility conditions.
- *reproducibility conditions*: Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

In the context of forensic tool test as described in [Gray12, p. 5] this amounts to *repeatability* being defined as the ability to get the same test results on the same testing environment (e.g. same computer, disk, mode of operation). *Reproducibility* is defined as the ability to get the same test results on a different testing environment (e.g. different PC, hard disk, operator).

While the forensic tool testing as described in [Gray12, p. 5] addresses exclusively digital forensics, first attempts to establish a benchmarking for digitised forensics are outlined in the following.

10.6.2 Benchmarking for digitised forensics on the example of digitised forensic dactyloscopy

In this section the exemplary chosen approach concerning a benchmarking in digitised forensics, namely the digitised forensic dactyloscopy, is outlined. As described in [KLD+11, p. 78810G-1], currently new contact-less optical fingerprint sensors are designed and developed to capture latent fingerprints without the utilisation of any development physical or chemical based technique, which can have an impact on follow up examinations of the fingerprint residue or render impossible (see also Sections 2.8 and 7.1). However, the appropriateness and the capabilities need the same amount of testing with as much rigor as the forensic tool testing for digital forensics from Section 10.6.1. The benchmarking as outlined in [KLD+11, p. 78810G-1] is even broader in scope. It includes items from the legal point of view and also includes the surfaces the contact-less sensor device is tested with. In [HML+11, p. 2] two important aspects are separated, which can be used to describe a contact-less benchmarking, properties from the forensic point of view, i.e. *what the device is capable of*, and from a technical point of view, i.e. *how the device works*. Those two groups are further divided. The forensic point of view comprises of legal requirements

properties *L*, Application related properties *A* and surface material properties *M*. The technical point of view contains technical properties *T*, input sensory properties *I* and processing methods *P*. The following Table 87 summarises the properties and supplies examples of sub-properties.

Property	**Sub properties (examples)**
Legal requirements properties *L*	Authenticity, integrity, privacy, evidentiary value, repeatability, documentation, general data protection
Application related properties *A*	Required pre-processing time, fingerprint detection performance, template matching performance, separation of overlapping fingerprints, age detection
Surface material properties *M*	Main material surface characteristics (surface finish, absorptive, structure pattern, deformability, shape), dimension of the sample object, main material, composition, substances on the surface: fingerprint (presence, completeness, overlapping, age), additional substances (presence, position relative to the fingerprint, type)
Technical properties *T*	Tolerated environmental factors, spatial resolution, scanning speed, scanning area, dimensions and transportability, maximum size of the scanned object
Input sensory properties *I*	Acquisition space, measurement method, mode of operation, frequency range, perspective distortion correction
Processing methods *P*	Pre-processing algorithms (e.g. Gabor filter), statistical measures (e.g. entropy, mean, standard deviation), subjective assessment (mostly research), differential imaging (mostly research)

Table 87: Properties and exemplary sub-properties in the benchmarking of contact-less fingerprint examination, modified from [HML+11, p. 2]

Using the instrument of the tuple to formalise a formal context, the properties can be connected as follows:

$$S = \{[L,A,M],[T,I,P]\} \qquad (10.6.2.1)$$

Here the properties in the first group of square brackets represent the forensic point of view, whereas the second group represents the properties comprising the technical point of view. Each of these properties comprise of sub-properties, which themselves can also contain sub-sub-properties. This leads to an extensible hierarchical structure, which can be supplemented if new benchmarking properties come to light.

In Section 4.4 it is shown, how benchmarking is integrated into the Data-Centric Examination Approach (DCEA), which is introduced in this thesis.

11. Bibliography:

[Ack89] R.L. Ackoff, “From data to wisdom”, Journal of Applied Systems Analysis16, pp 3–9, 1989

[ACPO20] Association of Chief Police Officers, "Good Practice Guide for Computer-Based Electronic Evidence", Official release version, [Online] https://www.7safe.com/docs/default-source/default-document-library/acpo_guidelines_computer_evidence_v4_web.pdf (06/03/2020)

[Ada19] Douglas Adams, "Is there an Artificial God?" [Online] http://www.biota.org/people/douglasadams/ (27/10/2019)

[AHK+19] Robert Altschaffel, Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, "Digital Forensics in Industrial Control Systems", In Proceedings of 38th International Conference of Computer Safety, Reliability, and Security (Safecomp 2019), Turku, Finland, September 11–13, Springer Nature Switzerland, ISBN 978-3-030-26600-4 pp. 128-136, 2019

[AKD09] Robert Altschaffel, Stefan Kiltz, Jana Dittmann, ”From the Computer Incident Taxonomy to a Computer Forensic Examination Taxonomy”, In 5th International Conference on IT Security Incident Management and IT Forensics, IMF 2009, Stuttgart, Germany, IEEE Computer Society, ISBN 978-0-7695-3807-5, DOI= 10.1109/IMF.2009.17, pp. 54-68, 2009

[AKD+15] Christian Arndt, Stefan Kiltz, Jana Dittmann, Robert Fischer, "ForeMan, a Versatile and Extensible Database System for Digitized Forensics Based on Benchmarking Properties", In Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'15), 17-19 June 2015, Portland, USA, pp. 91-96, ACM, DOI=10.1145/2756601.2756615, 2015

[ALK+18] Robert Altschaffel, Kevin Lamshöft, Stefan Kiltz, Mario Hildebrandt, Jana Dittmann, "A Survey on Open Forensics in Embedded Systems of Systems", In International Journal on Advances in Security, Vol. 11 No. 1 and 2, pp. 104-117, ISSN 1942-2636, 2018

[AnK97] Ross J. Anderson, Markus G. Kuhn, "Low Cost Attacks on Tamper Resistant Devices", In Proceedings of the 5th International Workshop on Security Protocols, Paris, France, Springer London, ISBN 3-540-64040-1, pp. 125-136, 1997

[Bar11] Jeffery G. Barnes, "History", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 1-7 - 1-22, 2011

[Bas19] Nanni Bassetti, "CAINE Live USB/DVD - computer forensics digital forensics"[Online] https://www.caine-live.net/ (01/09/2019)

[BeB10] Graeme B. Bell, Richard Boddington, “Solid State Drives: The Beginning of the End for Current Practice in Digital Forensic Recovery?”, In Journal of Digital Forensics, Security and Law, Vol 5. No. 3, pp. 1-20, 2010

[Bebi12] George Bebis, "CS479/679 Pattern Recognition - Introduction - Spring 2006" [Online] http://www.cse.unr.edu/~bebis/CS679/Lectures/Introduction.ppt (04/08/2012)

[BeC05] Nicola Lang Beebe, Jan Guyness Clark, “A Hierarchical, Objectives-Based Framework for the Digital Investigations Process”, In Digital Investigation Volume 2, Issue 2, pp. 146-166, 2005

[Bed20] Vladimir Bedecs, "SecureCam download" [online] https://sourceforge.net/projects/securecam/ (01/03/2020)

[Bee09] Nicola Beebe, “Digital Forensic Research: The Good, the Bad and the Unaddressed”, In Advances in Digital Forensics V, IFIP Advances in Information and Communication Technology 306, pp. 17-36, 2009

[BFGK09] Rainer Böhme, Felix C. Freiling, Thomas Gloe, Matthias Kirchner, "Multimedia Forensics Is Not Computer Forensics", In Proceedings of the 3rd International Workshop on Computational Forensics, IWCF09, Springer Verlag, ISBN 978-3-642-03520-3, DOI 10.1007/978-3-642-03521-0_9, pp. 90-103, 2009

[BHM08] Richard Boddington, Valerie Hobbs, Graham Mann, "Validating digital evidence for legal argument", In 6th Australian Digital Forensics Conference, 1-3 December, Perth, Australia, Edith Cowan University, pp. 1-17, 2008

[BKS13] Sebastian Breß, Stefan Kiltz, Martin Schäler, "Forensics on GPU coprocessing in databases - research challenges, first experiments, and countermeasures", In 15. Fachtagung des GI-Fachbereichs Datenbanken und Informationssysteme" (DBIS) des Business, Technologie und Web Workshops (BTW Workshops 2013), Magdeburg, Germany, 11-15.03.2013, pp 115-129, GI, ISBN 978-3-88579-610-7, 2013

[Bos07] Robert Bosch GmbH, "Bosch Automotive Electrics and Automotive Electronics" Springer Vieweg, ISBN 978-3-658-01783-5, DOI10.1007/978-3-658-01784-2, 2007

[BSI20] Bundesamt für Sicherheit in der Informationstechnik, "BSI-Standard 200-2 IT-Grundschutz Methodology" [online] https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/International/bsi-standard-2002_en_pdf (10/04/2020)

[BSI20a] Bundesamt für Sicherheit in der Informationstechnik, "BSI - Leitfaden IT-Forensik" [Online] https://www.bsi.bund.de/DE/Themen/Cyber-Sicherheit/Dienstleistungen/IT-Forensik/forensik_node.html (10/09/2020)

[Bur74] Arthur W. Burks, "Models of deterministic systems", In Journal of Mathematical systems theory, Vol.8, Issue 4, pp. 295-308, DOI: 10.1007/BF01780577, 1974

[BuW06] Steve Bunting, Wiliam Wei, “EnCase Computer Forensics: The Official EnCE: EnCase Certified Examiner Study Guide”, Wiley Publishing Inc., Indianapolis, USA, ISBN 978-0-7821-4435-2, 2006

[CaG04] Brian Carrier, Joe Grand, "A Hardware-Based Memory Acquisition Procedure for Digital Investigations" In Digital Investigation, Vol. 1, No. 1., doi: 10.1016/j.diin.2003.12.001, pp. 50-60, 2004

[CAP+10] Roberto Caldell, Irene Amerini, Francesco Picchioni, Alessia De Rosa, Francesca Uccheddu, "Multimedia Forensic Techniques for Acquisition Device Identification and Digital Image Authentication", In Handbook of Research on Computational Forensics, Digital Crime, and Investigation: Methods and Solutions, ISBN 9781605668369, pp. 130-151, 2010

[Car03] Brian Carrier, “Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers”, In International Journal of Digital Evidence, Volume 1, Issue 4, pp. 1-12, 2003

[Car03a] Brian Carrier, "Open Source Digital Forensics Tools - The Legal Argument", Tech Report, September 2003 [Online] http://www.digital-evidence.org/papers/opensrc_legal.pdf (31/10/2012)

[Car05] Brian Carrier, "File System Forensic Analysis", Addison Wesley Professional, ISBN 0-32-126817-2, 2005

[Car06] Brian D. Carrier, “A hypothesis-based approach to digital forensic investigations”, PhD Thesis, Purdue University West Lafayette USA, ISBN 978-0-542-86437-7, 2006

[Cas02] Eoghan Casey, “Error, Uncertainty and Loss in Digital Evidence”, In International Journal of Digital Evidence, Volume 1, Issue 2, pp. 1-45 , 2002

[Cas11] Eoghan Casey, “Digital evidence and computer crime: forensic science, computers and the Internet”, Academic Press, Elsevier Inc., ISBN 978-0-12--374268-1, 2011

[Cas19] Eoghan Casey, "Standardization of forming and expressing preliminary evaluative opinions on digital evidence", In Forensic Science International: Digital Investigation, Vol 32, Elsevier Ltd., DOI: 10.1016/j.fsidi.2019.200888, 2019

[CaS03] Brian Carrier, Eugene H. Spafford, ”Getting Physical with the Digital Investigation Process”, In International Journal of Digital Evidence, pp. 1-20, Volume 2, Issue 2, 2003

[CaS05] Brian D. Carrier, Eugene H. Spafford, "Automated digital evidence target definition using outlier analysis and existing evidence", In Refereed proceedings of the 5th Annual Digital Forensic Research Workshop (DFRWS), New Orleans, Louisiana, USA, August 17-19, pp. 1-10, 2005+

[CaS20] Eoghan Casey, Thomas R. Souvignet, "Digital transformation risk management in forensic science laboratories", In Forensic Science International Vol. 316, Elsevier B.V., DOI: 10.1016/j.forsciint.2020.110486, 2020

[CBB15] Eoghan Casey, Greg Back, Sean Barnum, "Leveraging CybOX™ to standardize representation and exchange of digital forensic information", In Digital Investigation, Elsevier Ltd, Vol. 12, DOI: 10.1016/j.diin.2015.01.014, 2015

[CBT20] Eoghan Casey, Maria Angela Biasiotti, Fabrizio Turchi, "Using Standardization and Ontology to Enhance Data Protection and Intelligent Analysis of Electronic Evidence" [Online] https://serval.unil.ch/resource/serval:BIB_EFAFD05944CB.P001/REF.pdf (18/09/2020)

[CFS20] J. Case, M. Fedor, M. Schoffstall, J. Davin, "A Simple Network Management Protocol (SNMP)" [Online] https://tools.ietf.org/html/rfc1157 (06/01/2020)

[CGS09] M. Cohen, S. Garfinkel, B. Schatz, "Extending the advanced forensic format to accommodate multiple data sources, logical evidence, arbitrary information and forensic workflow." DOI: 10.1016/j.diin.2009.06.010, In Digital Investigation, 6 (Supplement), ISSN:1742-2876, pp. 57-68, 2009

[ChT00] W.J. Chisum, B. Turvey, "Evidence Dynamics: Locard's Exchange Principle & Crime Reconstruction," In Journal of Behavioral Profiling, pp. 45-57, Volume 1, Issue 1, 2000

[CIN09] Committee on Identifying the Needs of the Forensic Sciences Community, "Strengthening Forensic Sciences in the United States: A Path Forward", US Department of Justice, ISBN 0-309-13131-6, 2009 [Online] http://www.nap.edu/catalog/12589.html (01/09/2014)

[Con12] John Convers, "Federal Rules of Evidence", The Committee on the Judiciary, House of Representatives; U.S. Government Printing Office: Washington, DC, United States of America, 1 December 2010 [Online] http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2010%20Rules/Evidence.pdf (01/11/2012)

[Cro07] Paul Crowley, "CD and DVD Forensics - Handle, Examine, and Process CD and DVD Evidence for Computer Forensics", Syngress Publishing Inc, ISBN 978-1-59749-128-0, 2007

[CVD+10] Ellick Chan, Shivaram Venkataraman, Francis David, Amey Chaugule, Roy Campbell, "Forenscope: A Framework for Live Forensics", Proceedings of the 26th Annual Computer Security Applications Conference ACSAC, Austin, Texas, USA, ACM, ISBN 978-1-4503-0133-6, pp. 307-316, DOI 10.1145/1920261.1920307, 2010

[Dar10] Glen S. Dardick, “Cyber Forensics Assurance”, In Proceedings of the 8th Australian Digital Forensics Conference, Perth, Australia, November 30th 2010, School of Computer and Information Science, Edith Cowan University, Perth, Western Australia, ISBN 978-0-7298-0685-5, pp. 57-64, 2010

[Del20] Guillaume Delugre, "Closer to metal: Reverse engineering the Broadcom NetExtreme's firmware" [online] https://web.archive.org/web/20110716102118/http://esec-lab.sogeti.com/dotclear/public/publications/10-hack.lu-nicreverse_slides.pdf (05/03/2020)

[Deri04] Luca Deri, "Improving Passive Packet Capture: Beyond Device Polling", In Proceedings of the 4th International System Administration and Network Engineering Conference (SANE2004), Amsterdam, The Netherlands, DOI: 10.1.1.58.3128, pp. 1-12, 2004

[Dew12] Andreas Dewald, "Formalsierung digitaler Spuren und ihre Einbettung in die forensische Informatik", Ph.D Thesis at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 2012

[Dew15] Andreas Dewald, "Characteristic Evidence, Counter Evidence and Reconstruction Problems in Forensic Computing", In Proceedings of the Ninth International Conference on IT-Security Incident Management and IT Forensics, Magdeburg, Germany, May 18th – 20th, 2015, IEEE Computer Society Conference Publishing Services (CPS), ISBN 978-1-4799-9903-3, pp. 77-82, DOI 10.1109/IMF.2015.15, 2015

[DHS01] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern classification", Wiley Interscience, New York, 2nd edition, ISBN 0-471-05669-3, 2001

[DiG01] L. Dixon, B. Gill, "Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision", monograph report, RAND Institute for Civil Justice, ISBN: 0-8330-3088-4, 2001

[Dil18] Ausra Dilijonaite, "Digital Forensic Readiness" In "Digital Forensics" edited by André Årnes, John Wiley and Sons Inc., ISBN 978-1119262381, 2018

[Dom19] Jeroen Domburg, "Sprites mods - Hard disk hacking" [Online] http://spritesmods.com/?art=hddhack (10/10/2019)

[Don20] Marco De Donno, "About PiAnoS — PiAnoS - Picture Annotation System dev documentation" [Online] https://ips-labs.unil.ch/doc/about.html (20/02/2020)

[Ecl20] eclypsium Inc., "Perilous peripherals: The Hidden Dangers Inside Windows and Linux Computers" [Online] https://eclypsium.com/2020/2/18/unsigned-peripheral-firmware/ (21/02/2020)

[EUC19] European Parliament and the Council, "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)" [Online] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN, OJ L 119, 4.5.2016, p. 1–88 (29/11/2019)

[Eur20] European Network of Forensic Science Institutes, "Guideline for Evaluative Reporting in Forensic Science" [Online] http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf (16/09/2020)

[FaV05] Dan Farmer, Wietse Venema, "Forensic Discovery", Addison-Wesley, ISBN 0-201-63497-X, 2005

[FCC+20] FCCID.io, "HIC-RJ0445 LS1 LITE-Xe" [Online] https://fccid.io/HIC-RJ0445 (25/03/2020)

[FBI14] The Federal Bureau of Investigation, "FBI-IAFIS" [Online] http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis (22/10/2014)

[FHP+12] Felix C. Freiling, Dirk Heckmann, Radim Polcak, Joachim Prosegga, "Forensic Computing (Dagstuhl Seminar 11401)", In Dagstuhl Reports, Volume 1 Number 10, pp. 1-13, DOI: 10.4230/DagRep.1.10.1, 2012

[Fie14] R. Fielding, "RFC1808: Relative Uniform Resource Locators", Standards Track, June 1995 [Online] http://www.w3.org/Addressing/rfc1808.txt (22/10/2014)

[For12] William Ford, "Forensic Information Data Exchange (FIDEX) Implementation Guide 2008-IJ-CX-K045", Tech Report, 05 May 2010, National Forensic Science Technology Center [Online] http://www.nfstc.org/wp-content/uploads/2011/06/Implementation-Guide_Final.pdf (03/06/2012)

[For14] Mark Forsyth, "The Unknown Unknown - Bookshops and the delight of not getting what you wanted", Icon Books Ltd. ISBN 978-184831-748-0, 2014

[FrS07] F. Freiling, B. Schwittay "A Common Process Model for Incident Response and Digital Forensics", In Proceedings of the IMF2007, Stuttgart, Germany, pp. 19-40, 2007.

[FRT20] FRT GmbH, "MicroProf200" [Online] https://frtmetrology.com/fileadmin/Daten/Images/02_Products/02_01_MicroProf/02_01_02_MicroProf200/MicroProf_200_ENG.pdf (25/03/2020)

[FSF20] Free Software Foundation, "Ddrescue - Data recovery tool" [online] http://www.gnu.org/software/ddrescue/ddrescue.html (09/03/2020)

[FT14] Forensic Technology, "IBIS: Integrated Ballistics Identification System | LDAS | RDAS", [Online] http://www.forensictechnology.com/heritage/ (20/10/2014)

[FVH+13] Robert Fischer, Claus Vielhauer, Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, "Ballistic examinations based on 3D data - a comparative study of probabilistic Hough Transform and geometrical shape determination for circle-detection on cartridge bottoms" In Proceedings of SPIE 8665 Media watermarking, security, and forensics 2013. - Burlingame, USA, 05-07.02.2013, pp. 86650F-86650F-12, SPIE, 2013

[Gar07] Simson Garfinkel, "Anti-Forensics: Techniques, Detection and Countermeasures", In 2nd International Conference on i-Warfare and Security, Naval Postgraduate School, Monterey, California, Academic Conferences Limited Reading UK pp. 77-84, 2007

[Gar10] Simson Garfinkel, "Digital Forensics Research: The next 10 years", In Digital Investigation, Issue 7, pp 64-73, ISSN 1742-2876, DOI 10.1016/2010.05.009, 2010

[GFC11] Miroslav Goljan, Jessica Fridrich, Mo Chen, "Defending Against Fingerprint-Copy Attack in Sensor-Based Camera Identification", In IEEE Transactions on Information Forensics and Security, Volume 6, Issue 1, pp. 227-236, IEEE Signal Processing Society, ISSN 1556-6013, DOI=10.1109/TIFS.2010.2099220, 2011

[Gla04] Pavel Gladyshev, "Formalising event reconstruction in digital investigations", PhD Dissertation, University College Dublin , 2004

[Gray12] Marty Gray, "NIST: General Test Methodology for Computer Forensic Tools - V 1.9", 07. November 2001 [Online] http://www.cftt.nist.gov/Test%20Methodology%207.doc (04/10/2012)

[GrR15] Jonathan Grier, Golden G. Richard III, "Rapid forensic imaging of large disks with sifting collectors", In Digital Investigation, Elsevier Ltd. Issue 2015, Volume 14, pp. S34-S445, DOI=10.1016/j.diin.2015.05.006, 2015

[Gut19] Peter Gutmann, "Secure Deletion of Data from Magnetic and Solid-State Memory" [Online] https://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html (29/10/2019)

[HaD04] Steve J. Harrison, Ronald H. Dieck, "Measurement Errors and Uncertainty", In The Engineering Handbook, Second Edition, CRC Press, ISBN 978-0-8493-1586-2, pp. 152.1-151.6, 2004

[Ham18] Jeff Hamm, "Computer Forensics" In "Digital Forensics" edited by André Årnes, John Wiley and Sons Inc., ISBN 978-1119262381, 2018

[Har71] J. Hartmanis, "Computational complexity of random access stored program machines", In Mathematical systems theory, ISSN 0025-5661, Springer-Verlag, Vol. 5, Issue 3, DOI: 10.1007/BF01694180 pp. 232-245, 1971

[HiD15] Mario Hildebrandt, Jana Dittmann, "StirTraceV2.0: Enhanced Benchmarking and Tuning of Printed Fingerprint Detection", In IEEE Transactions on Information Forensics and Security, Volume 10, Issue 4, DOI: 10.1109/TIFS.2015.2405412, pp. 833-848, 2015

[Hil20] Mario Hildebrandt, "On Digitized Forensics - Novel Acquisition and Analysis Techniques for Latent Fingerprints based on Signal Processing and Pattern Recognition", PhD Thesis, to be reviewed at Otto-von-Guericke-University, Faculty of Computer Science. 2020

[HKD08] Tobias Hoppe, Stefan Kiltz, Jana Dittmann, "Security threats to automotive CAN networks, practical examples and selected short-term countermeasures", In Computer Safety, Reliability, and Security, Proceedings of the 27th International Conference SAFECOMP 2008, Newcastle, UK, 22.09.-25.09.2008, Springer LNCS 5219, pp. 235-248, Michael D. Harrison and Mark-Alexander Sujan (Eds.), ISBN 978-3-540-87697-7, 2008

[HKD09] Tobias Hoppe, Stefan Kiltz, Jana Dittmann, "Applying Intrusion Detection to Automotive IT – Early Insights and Remaining Challenges" In Journal of Information Assurance and Security (JIAS), Volume 4, Issue 3, pp. 226-235, Dynamic Publishers, ISSN 1554-1010, 2009

[HKD11] Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, ”A Common Scheme for Evaluation of Forensic Software“, Sixth International Conference on IT Security Incident Management and IT Forensics (IMF), Stuttgart, Germany, IEEE Computer Society, ISBN 978-1-4577-0146-7, DOI 10.1109/IMF.2011.11, pp.92 - 106, 2011

[HKD11a] Tobias Hoppe, Stefan Kiltz, Jana Dittmann, "Security threats to automotive CAN networks - Practical examples and selected short-term countermeasures" In Journal of Reliability Engineering and System Safety, Volume 96, Issue 1, pp. 11-25, Elsevier Science, ISSN 0951-8320, DOI=10.1016/j.ress.2010.06.026, 2011

[HKD13] Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, "Digitized forensics - retaining a link between physical and digital crime scene traces using QR-codes", In Proceedings of SPIE 8667 Multimedia content and mobile devices 2013, Burlingame, USA, 03-07.02.2013, pp 86670S-86670S-11, SPIE, 2013

[HKD+14] Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, Claus Vielhauer,"An enhanced feature set for pattern recognition based contrast enhancement of contact-less captured latent fingerprints in digitized crime scene forensics", In Proc. SPIE 9028, Media Watermarking, Security, and Forensics 2014, San Francisco, USA, 02.02.2014, pp 902808-902808-15, SPIE, 2014

[HKG+11] Mario Hildebrandt, Stefan Kiltz, Ina Grossmann, Claus Vielhauer, "Convergence of digital and traditional forensic disciplines: a first exemplary study for digital dactyloscopy", In Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security (MM&Sec '11) Buffalo New York, USA, 29.09-30.09.2011, DOI: 10.1145/2037252.2037254, Publisher: ACM, pp. 1-8, 2011.

[HML+11] Mario Hildebrandt, Ronny Merkel, Marcus Leich, Stefan Kiltz, Jana Dittmann, Claus Vielhauer, "Benchmarking contact-Less surface measurement devices for fingerprint acquisition in forensic investigations: results for a differential scan approach with a chromatic white light sensor" In Proceedings of 17th International Conference on Digital Signal Processing (DSP), July 6th-8th Piscataway New Jersey USA, IEEE, ISBN 978-1-457-70273-0, pp. 1-6, DOI=10.1109/ICDSP.2011.6004969, 2011

[HMZ12] P. Hoffman, L. Masinter, J. Zawinski, "RFC2368: The mailto URL scheme", Standards Track, July 1998 [Online] http://www.ietf.org/rfc/rfc2368.txt (14/08/2012)

[Hof20] Heinz Hofbauer, "Visual Quality Index Implementations (VQI)" [Online] http://www.wavelab.at/sources/VQI/ (25/03/2020)

[HoL12] John D. Howard, Thomas A. Longstaff, "A Common Language for Computer Security Incidents", In Sandia National Laboratories Tech. Report SAND98-8667, 1998 [Online] http://infoserve.sandia.gov/sand_doc/1998/988667.pdf (06/10/2016)

[Hör20] Maël Hörz, "HxD - Freeware Hex Editor and Disk Editor" [online] https://mh-nexus.de/en/hxd/ (01/03/2020)

[IEEE98] Institute of Electrical and Electronics Engineers Inc. (IEEE), "IEEE Standard for Information Technology - Requirements and Guidelines for Test Methods Specifications and Test Method Implementations for Measuring Conformance to POSIX(R) Standards", ISO/IEC 13210-1999 IEEE Std. 2003-1997, DOI=10.1109/IEEESTD.1998.89201, 1998

[InR01] Keith Inman and Norah Rudin, “Principles and Practises of Criminalistics: The Profession of Forensic Science”, CRC Press LLC Boca Raton Florida, USA, ISBN 0-8493-8127-4, 2001

[ISO19] International Organization for Standardization (ISO), "ISO/SAE CD 21434 Road Vehicles -- Cybersecurity engineering" [Online] https://www.iso.org/standard/70918.html (13/09/2019)

[ISO94] International Organization for Standardization (ISO), "Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1: General Principles and Definitions", ISO 5725-1:1994, Technical Corrigendum 1, Published 1998-02-15, Geneva, Switzerland, 17p, 1994

[ISO94a] International Organization for Standardization (ISO), "Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method", ISO 5725-2:1994, First Edition 1994-12-15, Geneva, Switzerland, 42p, 1994

[ISO99] International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC), "General Requirements for the Competence of Testing and Calibration Laboratories", ISO/IEC 17025:1999, Geneva Switzerland, 26p, 1999

[ISO19] International Organization for Standardization, "ISO - ISO 7498:1984 - Information processing systems -- Open Systems Interconnection -- Basic Reference Model" [Online] https://www.iso.org/standard/14252.html (07/09/2019)

[JaA12] Wayne Jansen, Rick Ayers, "Guidelines on Cell Phone Forensic", Recommendations of the National Institute of Standards and Technology, Special Publication 800-101 [Online] http://csrc.nist.gov/publications/nistpubs/800-101/SP800-101.pdf (12/08/2012)

[JJNN08] A.K. Jain, Feng Jianjiang, A. Nagar, K Nandakumar, "On matching latent fingerprints", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (CVPRW '08, 23-28 June 2008), doi: 10.1109/CVPRW.2008.4563117, pp.1-8, 2008

[JCD07] Anil K. Jain, Yi Chen, Meltem Demirkus “Pores and Ridges: High-Resolution Fingerprint Matching Using Level 3 Features”, In IEEE Trans. Pattern Anal. Mach. Intell., ISSN 0162-8828, pp 15-27, 2007

[Jor19] Jason Jordaan, "The GDPR and DFIR" [Online] https://www.sans.org/cyber-security-summit/archives/file/summit_archive_1513005472.pdf (30/12/2019)

[JoS11] R. C. Joshi, Anjali Sardana, “Honeypots - A New Paradigm to Information Security”, CRC Press, Taylor and Francis Group, Boca Raton Florida, USA, ISBN 978-1-57808-708-2, 2011

[Kab12] M. E. Kabay, "The Parkerian Hexad", Norwich University, USA Course Material Seminar 1 Week 1 [Online] http://www.mekabay.com/overviews/hexad_ppt.zip (17/09/2012)

[KaK20] Corey Kallenberg, Xeno Kova, "How Many Million BIOSes Would you Like to Infect?" [online] https://legbacore.com/Research_files/HowManyMillionBIOSWouldYouLikeToInfect_Full2.pdf (02/03/2020)

[Kas19] Kaspersky Lab HQ, "Equation Group: Questions and Answers" [Online] https://cdn1.vox-cdn.com/uploads/chorus_asset/file/3415904/ Equation_group_questions_and_answers.0.pdf (27/10/2019)

[KDV12] Stefan Kiltz, Jana Dittmann, Claus Vielhauer, „Beweissichere Daten in der digitalisierten Forensik“, In D-A-CH Security 2012- Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven, Konstanz, Germany 25-26.09.2012, IT Verlag Sauerlach, ISBN 978-3-00-039221-4, pp. 288-300, 2012

[KDV15] Stefan Kiltz, Jana Dittmann, Claus Vielhauer "Supporting Forensic Design - a Course Profile to Teach Forensics", In Proceedings of the Ninth International Conference on IT-Security Incident Management and IT Forensics, Magdeburg, Germany, May 18th – 20th, 2015, IEEE Computer Society Conference Publishing Services (CPS), ISBN 978-1-4799-9903-3, pp. 85-95, DOI 10.1109/IMF.2015.15, 2015

[KCD+13] Stefan Kiltz, Eric Clausing, Jana Dittmann, Claus Vielhauer, „Ein Vorgehensmodell für die digitale Schlossforensik“, In D-A-CH Security 2013- Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven, Nürnberg, Germany, 17-18.09.2013, syssec, pp. 367-379, ISBN 978-3-00-042097-9, 2013

[KCG+12] Karen Kent, Suzanne Chevalier, Tim Grance, Hung Dang, “Guide to Integrating Forensic Techniques into Incident Response”, National Institute for Standards and Technology, NIST Special Publication 800-86 [Online] http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf (24/02/2012)

[KH02] W.G. Kruse and J.G. Heiser “Computer Forensics - Incident Response Essentials” Addison & Wesley, ISBN 0-201-70719-5, 2002

[KHA+09] Stefan Kiltz, Mario Hildebrandt, Robert Altschaffel, Jana Dittmann, Claus Vielhauer, Carsten Schulz, „Sicherstellung von gelöschtem Schadcode anhand von RAM-Analysen und Filecarving mit Hilfe eines forensischen Datenmodells“, In Sichere Wege in der vernetzten Welt: Tagungsband zum 11. Deutschen IT-Sicherheitskongress, Bonn, Deutschland, 12.05.-14.05.2009, SecuMedia Verlag Ingelheim, ISBN 978-3-922746-97-3, S. 473-488, 2009

[KHA+10] Stefan Kiltz, Mario Hildebrandt, Robert Altschaffel, Jana Dittmann, “A transparent bridge for forensic sound network traffic data acquisition“, In Sicherheit 2010 - Sicherheit, Schutz und Zuverlässigkeit, 5. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI) Berlin, Germany, 5-7 October 2010, pp. 93-104, 2010

[KHD09] Stefan Kiltz, Tobias Hoppe, Jana Dittmann, "A new forensic model and its application to the collection, extraction and long term storage of screen content off a memory dump", In Proceedings of the 16th International Conference on Digital Signal Processing (DSP'09) Santorini, Greece, IEEE Press, ISBN 978-1-4244-3297-4, pp. 1135-1140, 2009

[KHD+09] Stefan Kiltz, Tobias Hoppe, Jana Dittmann, Claus Vielhauer, "Video surveillance: The forensically sound retrieval and investigation of picture content off a memory dump", In Informatik 2009 Jahrestagung der Gesellschaft für Informatik e.V. (GI) Lübeck, Germany, 28.09.-02.10.2009, Gesellschaft für Informatik e.V. (GI) ISBN 978-3-88579-248-2, pp.174-189, 2009

[KHD+12] Stefan Kiltz, Mario Hildebrandt, Jana Dittmann, Claus Vielhauer, "Challenges in contact-less latent fingerprint processing in crime scenes: review of sensors and image processing investigations", In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucarest, Romania, 27-31.08.2012, pp. 1504-1508, 2012

[KHDV12] Stefan Kiltz, Mario Hildebrandt, Jana Dittmann, Claus Vielhauer, "Challenges in Contact-Less Latent Fingerprint Processes in Crime Scenes: Review of Sensors and Image Processing Investigations", In Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), 27-31 August 2012, Bucarest, Romania, European Association for Signal Processing, ISSN 2076-1465, pp. 1504-1508, 2012

[KLD+11] Stefan Kiltz, Marcus Leich, Jana Dittmann, Claus Vielhauer, Michael Ulrich, "Revised Benchmarking of Contact-less Fingerprint Scanners for Forensic Fingerprint Detection: Challenges and Results for Chromatic White Light Scanners (CWL)", In Proceedings of SPIE-IS&T Electronic Imaging, Vol. 7881A, Multimedia on mobile devices 2011, Bellingham, Washington, USA, Society of Photo-Optical Instrumentation Engineers (SPIE), ISBN 978-0-8194-8418-5, DOI=10.1117/12.872362, pp. 78810G-78810G-15, 2011

[KLD08] Stefan Kiltz, Andreas Lang, Jana Dittmann, "Taxonomy for Computer Security Incidents", In Cyber Warfare and Cyber Terrorism, Publisher Information Science Reference (IGI Global), Editors Lech J. Janczewski and Andrew M. Colarik, ISBN 978-1-59140-991-5, pp. 412-417, 2008

[Krae13] Christian Krätzer, "Statistical Pattern Recognition for Audio Forensics - Empirical Investigations on the Application Scenarios Audio Steganalysis and Microphone Forensics", PhD Thesis, Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Germany, Mai, 2013

[Kos14] S. Kostecke, "ntp.org: Home of the Network Time Protocol" [Online] http://www.ntp.org/ (20/10/2014)

[KVL11] Tobias Kiertscher, Claus Vielhauer, Marcus Leich, "Automated Forensic Fingerprint Analysis: A Novel Generic Process Model and Container Format", In: C. Vielhauer et al. (Eds) Biometrics and ID Management, Kongress: BioID (Brandenburg) 2011.03.08-10, Springer, Berlin, ISBN 3-642-19529-6, pp. 262-273, Lecture notes in computer science 6583, DOI=10.1007/978-3-642-19530-3_24, 2011

[Lah98] Bhagawandas Pannalal Lahti, "Signal Processing and Linear Systems", Berkeley-Cambridge Press, California, USA, ISBN 0-941413-35-7, 1998

[Lan07] Andreas Lang, "Audio Watermarking Benchmarking – A Profile Based Approach", Ph.D. thesis, Dept. of Computer Science, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany, 2007

[Lau13] Thomas Laurenson, "Performance Analysis of File Carving Tools", In proceedings of Security and Privacy Protection in Information Processing Systems (SEC 2013), IFIP Advances in Information and Communication Technology Vol 405, Springer, ISBN 978-3-642-39217-7, DOI: 10.1007/978-3-642-39218-4_31, pp 419-433, 2013

[LBF+09] Dale Liu, James Burton, Tony Fowly, Paul A. Henry, Jan Kanclirz Jr., Dave Kleiman, Thomas Millar, Kevin O'Shea, James Steele, Scott Sweitzer, Craig Wright, "Cisco Router and Switch Forensics: Investigating and Analyzing Malicious Network Activity", Syngress Publishing Inc., ISBN 978-1-59749-418-2, 2009

[LCL+14] Michael Hale Ligh, Andrew Case, Jamie Levy, Aaron Walters, "The Art of Memory Forensics - Detecting Malware and Threats in Windows, Linux and Mac Memory", John Wiley & Sons, Inc., ISBN: 978-1-118-82509-9, 2014

[Lew04] Julie A. Lewis, "Where Data Resides - Data Discovery from the Inside Out", Digital Mountain Inc., [Online] http://digitalmountain.com/fullaccess/Article3.pdf, 2004, (20/10/2014)

[LiB06] Vincent Liu, Francis Brown, “Bleeding-Edge Anti-Forensics”, Infosec World Conference & Expo, Orlando, Florida, USA, MIS Training Institute, 2006

[LeL09] Brian Neil Levine, Marc Liberatore, "DEX: Digital evidence provenance supporting reproducibility and comparison", In Digital Investigation, Vol 6, Elsevier Ltd, DOI: 10.1016/j.diin.2009.06.011, pp. S48-56, 2009

[LGS+10] Terrence V. Lillard, Clint P. Garrison, Craig A. Schiller, James Steele, "Digital forensics for network, Internet, and cloud computing: a forensic evidence guide for moving targets and data", Syngress/Elsevier Inc., ISBN 978-1-59749-537-0, 2010

[LMS12] P. Leach, M. Mealling, R. Salz, "RFC 4122: A Universally Unique IDentifier (UUID) URN Namespace", Standards Track, July 2005 [Online] http://www.ietf.org/rfc/rfc4122.txt (2012)

[Loc20] Edmont Locard, "L'enquete criminelle et les methodes scientifique", Ernest Flammarion, Paris, 1920

[Lom03] M. A. Lombardi, "Radio Controlled Clocks", In Proceedings of the 2003 NCSL International Workshop and Symposium, Tampa, Florida, USA August 2003, p. 1-18, 2003

[LTP20] LabTestProject, "Download sha256sum.exe for Windows" [Online] http://www.labtestproject.com/files/sha256sum/sha256sum.exe (25/03/2020)

[Mac11] Alice V. Maceo, "Documentation of Friction Ridge Impressions: From the Scene to the Conclusion", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 10-3 - 10-20, 2011

[Mac11a] Alice V. Maceo, "Anatomy and Physiology of Adult Friction Ridge Skin", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 2-3 - 2-26, 2011

[MaLe70] Leonard A. Marascuilo, Joel R. Levin, "Appropriate Post Hoc Comparisons for Interaction and Nested Hypotheses in Analysis of Variance Designs: The Elimination of Type IV Errors", In American Educational Research Journal, Volume 7, Issue 3, pp. 397-421, American Educational Research Association, 1970

[Man19] Gavin W. Manes, et al., "Backup Tape Forensics is Here to Stay" [Online] https://www.avansic.com/News/Story/84/ (08/12/2019)

[MaS01] Douglas Mauro, Kevin Schmidt, "Essential SNMP" [Online] https://learning.oreilly.com/library/view/essential-snmp-2nd/0596008406/ch01.html (09/04/2020), O'Reilly & Associates Inc., ISBN 0-596-00020-0, 2001

[MaS01a] Douglas Mauro, Kevin Schmidt, "Essential SNMP" [Online] https://learning.oreilly.com/library/view/essential-snmp-2nd/0596008406/ch02.html (09/04/2020), O'Reilly & Associates Inc., ISBN 0-596-00020-0, 2001

[MBK+11] Ronny Merkel, Anja Bräutigam, Christian Krätzer, Jana Dittmann, Claus Vielhauer, "Evaluation of Binary Pixel Aging Curves of Latent Fingerprint Traces for Different Surfaces Using a Chromatic White Light (CWL) Sensor", In Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security, Buffalo, New York, USA, Association of Computers and Machinery, pp. 41-50, ISBN 978-1-4503-0806-9, DOI=10.1145/2037252.2037262, 2011

[Mer14] Ronny Merkel, "New solutions for an old challenge: chances and limitations of optical, non-invasive acquisition and digital processing techniques for the age estimation of latent fingerprints", PhD thesis, Otto von Guericke University Magdeburg, DOI: 10.25673/4091, pp. 1-214, 2014

[Meu15] Didier Meuwly, "Forensic Use of Fingerprints and Fingermarks", In Encyclopedia of Biometrics, Springer, IBSN: 978-1-4899-7488-4, DOI: 10.1007/978-1-4899-7488-4_181, pp. 723-735, 2015

[MeVe12] Didier Meuwly, Raymond Veldhuis, "Forensic biometrics: From two communities to one discipline", In Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG 2012), Darmstadt, Germany, ISSN 1617-5468, IEEE, pp. 1-12, 2012

[MiFe74] Ian I. Mitroff, Tom R. Featheringham, "On systemic problem solving and the error of the third kind", In Journal of Behavioral Science, Volume 19, Issue 6, pp. 383-393, Wiley & Sons Ltd., DOI=10.1002/bs.3830190605, 1974

[Mil96] Holmes Miller, "The Multiple Dimensions of Information Quality", In Journal of Information Systems Management, Volume 13, Issue 2, pp. 79-82, Taylor and Francis Group, DOI=10.1080/10580539608906992, 1996

[MoM11] Andre A. Moenssens, Stephen B. Meagher, "Fingerprints and the Law", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 13-3 - 13.26, 2011

[MöH19] Dietmar P. F. Möller, Roland E. Haas, "Guide to Automotive Connectivity and Cybersecurity", Springer International Publishing AG, ISBN 978-3-319-73511-5, https://doi.org/10.1007/978-3-319-73512-2, 2019

[Mus20] Tim Musson, "happy mailing : blat online" [online] https://www.blat.net/ (01/03/2020)

[NeR06] Ruprecht Nennstiehl, Joachim Rahm, "A Parameter Study Regarding the IBIS Correlator" In Journal of forensic sciences, Volume 51, Number 1, pp.18-23, DOI=10.1111/j.1556-4029.2005.00002.x, 2006

[New07] Robert C. Newman, “Computer Forensics: evidence collection and Management”, Auerbach Publications, Taylor and Francis Group, ISBN: 978-0-8493-0561-0, 2007

[Nie95] Hanne Riis Hielson, Flemming Nielson, "Semantics with Applications, A Formal Introduction", (1st ed.). Chicester, England, John Wiley & Sons., ISBN 0-471-92980-8. 1995

[Nik05] Bruce Nikkel, "Forensic acquisition and analysis of magnetic tapes" In Digital Investigation, Volume 2, DOI=10.1016/j.diin.2005.01.007, pp. 8-18, 2005

[NIST20] National Institute of Standards and Technology, "Computer Forensics Tool Testing Program (CFTT) | NIST" [Online] https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/disk (06/04/2020)

[NIST20a] National Institute of Standards and Technology, “Deleted File Recovery | NIST” [Online] https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/deleted (06/04/2020)

[NPMO14] NIEM Program Management Office "Introduction to the National Information Exchange Model (NIEM)", Tech Report, Document Version 0.3, 12 February 2007 [Online] https://www.niem.gov/documentsdb/Documents/Overview/NIEM_Introduction.pdf (20/10/2014)

[NYE17] Nir Nissim, Ran Yahalom, Yuval Elovici, "USB-based attacks", In Computers & Security, Vol. 70, DOI: 10.1016/j.cose.2017.08.002, pp. 675-688, 2017

[Ora19] Oracle Corporation, "Modifying Slices or Partitions - Managing Devices in Oracle® Solaris 11.2" [Online] https://docs.oracle.com/cd/E36784_01/html/E36834/gnvod.html (31/05/2019)

[Pal01] G. Palmer, “A Road Map for Digital Forensic Research”, DFRWS Technical Report, DTR-T001-01 Final, Air Force Research Laboratory, Rome, New York, 2001

[PaMi10] Joseph N. Pato, Lynette I. Millet, "Biometric Recognition: Challenges and Opportunities", National Academies of Sciences, ISBN 978-0-309-14207-6, National Academies Press, 2010

[Par97] Donn B. Parker, "Information Security in a Nutshell", In Journal of Information Systems Security, Volume 6, Issue 1, pp. 14-19, Taylor & Francis, DOI=10.1080/10658989709342524, 1997

[PBK+07] Sean Peisert, Matt Bishop, Sidney Karin, Keith Marzullo, "Towards models for forensic analysis", In Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering, pp. 3-15, 2007

[PBM08] Sean Peisert, Matt Bishop, Keith Marzullo, "Computer forensics in forensis", In SIGOPS Operating Systems Review, Volume 42, Issue 3, pp 112-122, ACM, DOI=10.1145/1368506.1368521, 2008

[PCJ+18] Mark Pollitt, Eoghan Casey, David-Olivier Jaquet-Chiffelle, Pavel Gladyshev, "A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence", In OSAC Technical Series 0002R1, DOI: 10.29325/OSAC.TS.0002, pp. 1-23, 2018

[Pol01] Mark M. Pollitt, "Report on Digital Evidence", In Proceedings of the 13th INTERPOL Forensic Science Symposium, Lyon, France, October 16-19, U.S. Department of Justice, pp. D4-85 - D4-134, 2001

[Pol08] Mark Pollitt, “Applying Traditional Forensic Taxonomy to Digial Forensics”, in IFIP International Federation for Information Processing, Volume 258; Advances in Digital Forensics IV, Indrajit Ray, Sujeet Shenoi, Boston: Springer, pp. 17-26, DOI: 10.1007/978-0-387-84927-0_2, 2008

[Pol10] Mark Pollitt, "A History of Digital Forensics", Advances in Digital Forensics VI, IFIP Advances in Information and Communication Technology, Volume 337/2010, DOI 10.1007/978-3-642-15506-2_1, pp. 3-15, 2010

[Pol20] Philip A. Polstra, "Bypassing Endpoint Security for $20 or less" [online] https://www.defcon.org/images/defcon-20/dc-20-presentations/Polstra/DEFCON-20-Polstra-Bypassing-Endpoint-Security.pdf (06/03/2020

[Pri20] Joshua Pritikin "Win32 native md5sum, sha1sum, sha256sum etc." [online] https://web.archive.org/web/20090613031908/http://blog.nfllab.com/archives/152-Win32-native-md5sum,-sha1sum,-sha256sum-etc..html (01/03/2020)

[Prw14] PRWeb, "Computer Card to Defeat Evidence in Cyber Security Analysis and Memory Forensics" [Online] http://www.prweb.com/releases/prweb2012/2/prweb9186495.htm (20/10/2014)

[PWF+06] Nick L. Petroni Jr., Aaron Walters, Timothy Fraser, William A. Arbaugh, "FATKit: A framework for the extraction and analysis of digital forensic data from volatile system memory", In Digital Investigation, Issue 3, pp. 197-210, DOI 10.1016/j.diin.2006.10.001, 2006

[Pyt20] The Python Foundation, "Welcome to python.org" [online] https://www.python.org/ (01/03/2020)

[QSS+12] Kun Qian, Maik Schott, Werner Schöne, Mario Hildebrandt, "Separation of High-Resolution Samples of Overlapping Latent Fingerprints Using Relaxation Labeling", In Proceedings of SPIE Photonics Europe, Volume 8436, 84361A, Optics, Photonics, and Digital Technologies for Multimedia Applications II, April 16th, 2012, Brussels, Belgium, Society of Photo-Optical Instrumentation Engineers (SPIE) pp. 84361A-84361A-9, DOI=10.1117/12.922696, 2012

[Rec06] Gerald Recktenwald, “Uncertainty Estimation and Calculation”, [Online] http://web.cecs.pdx.edu/~gerry/class/ME449/lectures/pdf/uncertaintySlides.pdf, 2006, (20/10/2014)

[Red19] Niranjan Reddy, "Practical Cyber Forensics An Incident-Based Approach to Forensic Investigations", Apress Berkeley CA, ISBN 978-1-4842-4459-3, 2019

[RGM+06] Marcus K Rogers, James Goldman, Rick Mislan, Timothy Wedge, “Computer Forensics Field Triage Process Model”, In Journal of Digital Forensics, Security and Law (JDFSL), Association of Digital Forensics, Security and Law (ADFSL), Volume 1, Number 2, pp. 27-40, 2006

[RiC14] Golden G. Richard III., Andrew Case, "In lieu of swap: Analyzing compressed RAM in Mac OS X and Linux", In Journal of Digital Investigation, Volume 11, Supplement 2, pp. S3-S11, Elsevier Science, ISSN 1742-2876, DOI=http://dx.doi.org/10.1016/j.diin.2014.05.011, 2014

[Ric19] Pat Richards, "Microchip: A CAN Physical Layer Discussion - ANN228" [Online] http://ww1.microchip.com/downloads/en/appnotes/00228a.pdf (15/10/2019)

[Ric20] Matthew Richardson, "PGP Digital Timestamping Service" [Online] http://www.itconsult.co.uk/stamper.htm (01/03/2020)

[Row07] Jennifer Rowley, “The wisdom hierarchy: representations of the DIKW hierarchy”, In Journal of Information Science, pp. 163-180, Volume 33, Issue 2, DOI: 10.1177/0165551506070706, 2007

[Rus19] Andrew L. Russell, "Osi The Internet That Wasnt - IEEE Spectrum" [Online] https://spectrum.ieee.org/tech-history/cyberspace/osi-the-internet-that-wasnt (07/09/2019)

[RXS+13] Huw Read, Konstantinos Xynos, Iain Sutherland, Gareth Davies, Tom Houiellebecq, Frode Roarson, Andrew Blyth, "Manipulation of hard drive firmware to conceal entire partitions", In Digital Investigation, Vol. 10, Issue 4, DOI: 10.1016/j.diin.2013.10.001, pp. 281-286, 2013

[RYG05] C. Ruibin, T. Chan Kai Yun, M. Gaertner, “Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework”, In International Journal of Digital Evidence, pp. 1-13, Volume 4, Issue 1, 2005

[SaG09] Antonio Savoldi, Paolo Gubian, "Blurriness in Live Forensics - An Introduction", In Advances in Information Security and its Application, Communications in Computer and Information Science, Volume 36, Part 3, DOI 10.1007/978-3-642-02633-1_16, pp. 119-126, 2009

[SaJ07] Tony Sammens and Brian Jenkinson, "Forensic computing: a practitioner's guide", Second Edition, Springer Verlag London-Berlin-Heidelberg, ISBN 978-1-84628-397-0, 2007

[Sch20] Patrick Schmidt, "Conclusion - 15 Years of Hard Drive History: Capacities Outran Performance" [Online] https://www.tomshardware.com/reviews/15-years-of-hard-drive-history,1368-11.html (22/02/2020)

[Sea19] Seagate Technology LLC., "SCSI Commands Reference Manual" [Online] https://www.seagate.com/files/staticfiles/support/docs/manual/Interface%20manuals/100293068j.pdf (22/12/2019)

[SEH03] Susan Elliot Sim, Steve Easterbrook, Richard C. Holt, "Using benchmarking to advance research: A challenge to software engineering", In Proceedings of the 25th International Conference on Software Engineering (ICSE 2003), IEEE Computer Society, pp. 74–83, DOI=10.1.1.13.1849, 2003

[Shi16] Goraksh Shinde, "SAP ERP Forensics", In International Journal for Advanced Research in Engineering and Technology (IJARET), Volume 4, Issue 1, p. 16-22, ISSN 2320-6802, 2016

[SiSh07] Howard Silverstone, Michael Sheetz, "Forensic Accounting and Fraud Investigation for Non-Experts", Wiley & Sons, ISBN 978-0-471-78487-6, 2007

[Ski20] Irfan Skiljan, "IrfanView - Official Homepage - One of the Most Popular Viewers Worldwide" [online] https://www.irfanview.com/ (01/03/2020)

[SoM14] K. Sollins, L. Masinter, "RFC1713: Functional Requirements for Uniform Resource Names", Status Informational, December 1994 [Online] http://www.ietf.org/rfc/rfc1737.txt (20/10/2014)

[Sri07] S. Srinivasan, "Security and Privacy vs. Computer Forensics Capabilities", In Information Systems Control Journal, Information Systems Audit and Control Association (ISACA), Volume 4, pp.1-3, 2007

[Sri10] Sargur N. Srihari, "Computing the scene of a crime", In IEEE Spectrum, Vol. 47, Issue 12, pp. 38-43, DOI: 10.1109/MSPEC.2010.5644777, 2010

[Str10] Gary Stringham, "Hardware/Firmware Interface Design - Best Practices for Improving Embedded Systems Development", ISBN: 978-1-85617-605-7, Elsevier Inc., 2010

[Sui20] Matthieu Suiche, "Win32dd is a free kernel land and 100% open-source tool to acquire physical memory" [online] https://web.archive.org/web/20090302050236/http://win32dd.msuiche.net/ (01/03/2020)

[TaK94] Barry N. Taylor, Chris E. Kuyatt, "NIST Technical Note 1297 1994 Edition - Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results", NIST Technical Note 1297, National Institute of Standards and Technology, Gaithersburg, Maryland, 1994

[Tan01] Andrew S. Tanenbaum, "Modern Operating Systems, 2nd Edition", Prentice Hall International, ISBN: 978-0130926418, 2001

[Tan01a] John Tan, "Forensic Readiness", Cambridge USA, @stake Inc., DOI=10.1.1.480.6094, 2001

[TGT20] The GIMP Team, "GIMP - GNU Image Manipulation Program" [online] https://www.gimp.org/ (13/04/2020)

[ThKo06] Sergios Theodoridis, Konstantinos Koutroumbas, "Pattern Recognition", In Academic Press, 3rd Edition, ISBN 0-12-369531-7, 2006

[Thw20] Doug Thwaites, "Find unknown devices using a vendor and device ID - Device Hunt" [online] https://devicehunt.com/ (01/03/2020)

[TIS93] TIS Committee, "Tool Interface Standard (TIS) - Formats Specification for Windows, Version 1.0", Technical Report, Intel Order Number 241597, [Online] http://www.x-ways.net/winhex/kb/ff/PE_EXE.pdf, 1993, (20/10/2014)

[TIS93a] TIS Committee, "Tool Interface Standard (TIS) - Portable Formats Specification, Version 1.1", Technical Report, Intel Order Number 241597, [Online] http://www.acm.uiuc.edu/sigops/rsrc/pfmt11.pdf, 1993 (20/10/2014)

[Tur05] Philip Turner, "Unification of Digital Evidence from Disparate Sources (Digital Evidence Bags)", In Journal of Digital Investigation, Volume 2, Issue 3, pp. 223-228, Elsevier Science, ISSN 1742-2876,DOI=10.1016/j.diin.2005.07.001, 2005

[Tur06] Philip Turner, "Selective and intelligent imaging using digital evidence bags", In Journal of Digital Investigation, Volume 3, Supplement, pp. 59-64, Elsevier Science, ISSN 1742-2876, DOI=10.1016/j.diin.2006.06.003, 2006

[UDR20] USBDev.ru, "Скачать программы для восстановления флешек" [online] http://www.usbdev.ru/files/ (08/03/2020)

[UDR20a] USBDev.ru, "ChipGenius v4.19.1225 (2019-12-25) by hit00" [online] http://www.usbdev.ru/files/chipgenius/ (08/03/2020)

[UDR20b] USBDev.ru, "UFDUtility v3.4.8.0" [online] https://www.usbdev.ru/files/ite_tech/ufdutility/ (08/03/2020)

[UDR20c] USBDev.ru, "Phison UPTool v2.094_20150909" [Online] http://www.usbdev.ru/files/phison/uptool/ (08/03/2020)

[UWa20] University of Waikato, "Weka - The workbench for machine learning" [Online] https://www.cs.waikato.ac.nz/ml/weka/ (25/03/2020)

[Van11] John R. Vanderkolk, "Examination Process", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 9-3 - 9.26, 2011

[Ver12] Erwin Vermeij, "Forensic Applications of Scanning Electron Microscopy, a brief overview" [Online] http://www.fei.com/uploadedFiles/Documents/Forensics/Forensic-Application-SEM.pdf (11/07/2012)

[Vie06] Claus Vielhauer, "Biometric User Authentication for IT-Security - From Fundamentals to Handwriting", In Advances in Information Security, Springer Science+Business Media Inc., ISBN 0-387-26194-X, 2006

[VMC12] Konstantinos Vlachopoulos, Emmanouil Magkos, Vassileios Chrissikopoulos, "A Model for Hybrid Evidence Investigation", In International Journal for Digital Crime Forensics, Vol 4, Issue 4, IGI Global, pp. 47-62, DOI=10.4018/jdcf.2012100104, 2012

[Vol20] Volatility Foundation, "The Volatility Foundation - Open Source Memory Forensics" [online] https://www.volatilityfoundation.org/ (01/03/2020)

[Wei02] M. C. Weil, "Dynamic Time & Date Stamp Analysis", In International Journal of Digital Evidence, Volume 1, Issue 2, pp. 1-6, 2002

[Wol09] Stephen D. Wolthusen, "Overcast: Forensic Discovery in Cloud Environments", In Proceedings of the 5th Conference on IT-Security Incident Management and IT Forensics, ISBN 978-0-76965-3807-5, 2009

[Wol20] Edward J. Wollack, "Seven Year Microwave Sky Image" [online] https://wmap.gsfc.nasa.gov/media/101080/101080_7yrFullSky_WMAP_1024W.png (26/03/2020)

[XWS20] X-Ways Software Technology AG, "WinHex: Computer Forensics & Data Recovery Software, Hex Editor & Disk Editor" [Online] (09/03/2020)

[XWS20a] X-Ways Software Technology AG, "X-Ways Forensics: Integrated Computer Forensics Software" [online] http://www.x-ways.net/forensics/ (10/03/2020)

[YaF11] Brian Yamashita, Mike French, "Latent print development", Chapter of "The Fingerprint Sourcebook", Eric H. Holder, Jr., Laurie O. Robinson, John H. Laub (eds), pp. 7-3 - 7-67, 2011

[YLJ15] Soweon Yoon, Eryun Liu, Anil K. Jain, "On Latent Fingerprint Image Quality", In International Workshop on Computational Forensics (IWCF 2014), Lecture Notes in Computer Science, Springer, Cham, pp 67-82, DOI=10.1007/978-3-319-20125-2_7, 2015

[ZHR+07] Jingmin Zhou, Mark Heckman, Brennen Reynolds, Adam Carlson, Matt Bishop, "Modelling Network Intrusion Detection Alerts for Correlation", In ACM Transactions on Information Systems Security, Volume 10, Issue 1, DOI = 10.1145/1210263.1210267, 2007