university of groningen

University Medical Center Groningen

University of Groningen

Reliability assessment of digital forensic investigations in the Norwegian police

Stoykova, Adi; Andersen, Stig; Franke, Katrin; Axelson, Stefan

Published in:
Forensic Science International: Digital Investigation

DOI:
10.1016/j.fsidi.2022.301351

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Stoykova, A., Andersen, S., Franke, K., & Axelson, S. (2022). Reliability assessment of digital forensic investigations in the Norwegian police. *Forensic Science International: Digital Investigation*, *40*, [301351]. https://doi.org/10.1016/j.fsidi.2022.301351

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-06-2022

Forensic Science International: Digital Investigation 40 (2022) 301351

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Reliability assessment of digital forensic investigations in the Norwegian police

Radina Stoykova [a, c, *], Stig Andersen [a, b, **], Katrin Franke [a], Stefan Axelsson [a, d]

[a] Norwegian University of Science and Technology, Teknologivegen 22, Gjøvik, 2815, Norway
[b] Oslo Police District, PO Box 2093 Vika, Oslo, 0125, Norway
[c] University of Groningen, PO Box 72, Groningen, 9700, AB, the Netherlands
[d] Stockholm University, DSV, Postbox 7003, Kista, 164 07, Sweden

ARTICLE INFO

Article history:
Received 12 October 2021
Received in revised form
20 January 2022
Accepted 23 January 2022
Available online xxx

Keywords:
Criminal investigation
Digital forensics
Reliability
Validation
Forensic reports

ABSTRACT

This case study presents a qualitative assessment of the reliability of digital forensic investigation in criminal cases in Norway. A reliability validation methodology based on international digital forensic standards was designed to assess to what extent those standards are implemented and followed by law enforcement in their casework. 124 reports related to the acquisition, examination, and analysis of three types of digital data sources - computers, mobile phones, and storage devices were examined. The reports were extracted from the criminal case management system used by the police and prosecution services. The reports were examined on technology, method, and application level in order to assess the reliability of digital evidence for criminal proceedings.

The study found that digital forensic investigation in 21 randomly sampled criminal cases in Norway were insufficiently documented to assess the reliability of the digital evidence. It was not possible to trace the digital forensic actions performed on each item or link the digital evidence to its source. None of the cases were shown to comply with digital forensic methodology, justify the methods and tools used, or validate tool results and error rates.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Norway has one of the highest human development value indices in the world with 96.5% of the population connected to the internet and only a homicide rate of 0.5 per 100 000 inhabitants (United Nations Development Programme, 2020). Cybercrime and the use of ICT in crime in general has increased in Norway for many years (Justis- og beredskapsdepartementet, 2012; Justis- og beredskapsdepartementet, 2020), and in a recent report, the Auditor general found that most of the available digital expertise in the Norwegian police is used to investigate digital evidence from serious crimes such as homicide, sexual assault, and narcotics (Riksrevisjonen, 2021). With digital evidence being of increasing importance in criminal investigation, law enforcement authorities are dependent on digital forensics methods to ensure evidence reliability and compliance with fair trial requirements of evidence authenticity, reliability, and contestability in criminal proceedings (Council of Europe Guide, 2021).

There is insufficient research on the extent to which law enforcement uses digital forensics methodology and tools, how they implement digital forensics guidelines and standards, and how digital data sources are acquired, examined, and analysed. Little is known about how investigative reports preserve chain of custody information in order to audit the digital forensic examination performed in each case. This study aimed to present the state of the art in digital evidence management and the use of digital forensics in Norwegian police and to anticipate the challenges in practice.

The study analysed data from the Norwegian crime investigation case management system in which all criminal cases, including all case documentation, is processed. The data set consist of reports from 21 homicide and sexual assault cases which lead to an indictment. The reports include information about the seized digital data sources and the digital forensic methods and tools used to acquire, examine, and analyse digital evidence. The scope of the

* Corresponding author. Norwegian University of Science and Technology, Teknologivegen 22, Gjøvik, 2815, Norway.
** Corresponding author. Norwegian University of Science and Technology, Teknologivegen 22, Gjøvik, 2815, Norway.
E-mail addresses: radina.r.stoykova@ntnu.no (R. Stoykova), stig.andersen@politiet.no (S. Andersen).

https://doi.org/10.1016/j.fsidi.2022.301351
2666-2817/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

study was limited to three types of data sources: computers, mobile phones, and storage devices. A total of 124 reports concerning 187 such devices were assessed. From each case, all reports concerning acquisition, examination and analysis of data from each of the three device types were extracted. Each stage was qualitatively evaluated in order to assess the reliability of the digital forensics methods, tools, and examiner work.

The scope of this study was the digital forensics investigation in the narrow sense, referring only to the scientific methodology employed to address a forensic task. The entire criminal investigation has the objective to discover, study, and explain facts used to inform decisions about incidents which violate criminal law (Innes, 2003; Stelfox, 2009; Brodeur, 2010) and includes investigative objectives beyond digital forensics investigation. The current proliferation of digital investigation process models for different purposes and technologies is best examined in the work of Kohn et al. (2013) and Montasari et al. (2015). However, there is a consensus in the digital forensic community that the scientific methodology employed in digital investigation is a separate stage in the investigation which at its minimum contains the following sub-stages: *acquisition*, *examination*, *analysis* and *reporting*. These stages of the investigation are described as *Digital forensic investigation* by Kohn, *In lab process* by Montasari, and *Laboratory analysis procedure* by Interpol (Global guidelines, 2019). We adopt the term digital forensic investigation (DFI) defined by Kohn as 'a special type of investigation where the scientific procedures and techniques used will allow the results, i.e., digital evidence, to be admissible in a court of law' (Kohn et al., 2013, p.104). This study was limited to the minimum digital forensics process - acquisition, examination, and analysis. The expectation was that these core stages were readily identifiable from the reports concerning digital forensics tasks performed in each case. The study, therefore, does not address other stages of the digital investigation such as preparation, incident response, crime scene investigation, etc. Reporting is addressed implicitly through the study itself rather than as an explicit process step.

This paper is organized as follows: Section 2 outlines previous studies on the use of digital forensics by law enforcement. It also discusses the digital forensic process in relation to selected international standards and a reliability validation framework as a background for the method in this study. Section 3 describes the method used in the study, including the reliability criteria and evaluation requirements for each of the three core process steps. The results are presented in section 4 and discussed in section 5 to outline further recommendations. The paper concludes with some suggestions for future research in section 6.

2. Background

The literature review in this section first presents previous research on the use of digital forensics in law enforcement criminal investigation work, in order to identify challenges. Second, the digital forensics investigation and process stages are described briefly, taking into account international standards, guidelines and best practices for reliability assurance. Finally, the framework for reliability assessment which is the background for the method of the study is discussed.

2.1. The use of digital forensics by law enforcement

Searches for academic work examining how law enforcement employ digital forensics (DF) and technology in their day-to-day work, and how the reliability and efficiency of such endeavours are assessed, yield few results. Only two quantitative studies of police data were identified. Statistical data from an Australian law enforcement agency was analysed to point to the increased use of digital evidence in relation to different types of crime (Turnbull et al., 2009). Further, a case study of the Dubai police discussed the increased volume and complexity of data which causes significant delays in investigations (Alawadhi et al., 2015). Two other studies focused on efficiency showing that the police often fails to effectively or strategically use investigative technology, stemming from financial issues and a lack of quality evaluation of both the technology and how law enforcement uses it (Koper et al., 2014; Custers and Vergouw, 2015). Studies in Norway suggest that the personnel examining and utilizing digital evidence do not possess the necessary competency to perform such tasks (Heitmann, 2019; Erlandsen, 2019) and that digital forensics examiners are prone to contextual bias and technology dependencies (Sunde and Dror, 2021).

Law enforcement must be able to demonstrate that the methodology and tools employed in the processing of digital data for evidence purposes are reliable and forensically sound. This means that the methodology must allow for reproducibility and validation (Risinger, 2018), and lead to consistent intended behavior and results (ISO/IEC 27037:2012). Validation is the scientifically accepted methodology for demonstrating the accuracy and reliability of a process (Hughes and Karabiyik, 2020). In digital forensics, the same initial conditions and input should always generate the same results. Validation depends not only on objective measurements e.g. algorithms, but also on subjective measures such as parameterisation of the method or tool by the examiner (Stoykova and Franke, 2020).

In recent years, digital forensics specialist and academics have appealed for the importance of reliability validation in digital forensics (Casey, 2019; Jones and Vidalis, 2019; Hughes and Karabiyik, 2020; Horsman, 2018a). Tully argues that reliability validation is currently reduced in practice—it is focused on tool verification and not overall method validation, main uncertainties are not identified, and there is a 'lack of evidence to demonstrate that the method is repeatable within units.' (Tully et al., 2020, p.5) Most digital forensics techniques have not satisfied the criteria of known error rates and a lack of resources and data sets for testing is identified (Jones and Vidalis, 2019). Therefore, it is necessary for law enforcement agencies to test the reliability of the digital forensics methods and tools they employ.

A gap in the literature on digital forensic reliability is that most studies describe high-level challenges or validation requirements, but they do not propose practical and implementable validation methodology which can be adopted by law enforcement in their daily work and standard operating procedures. This study develops such a practical reliability framework based on minimum documentation requirements for the methods, tools, and the interaction of examiners across each stage of digital evidence acquisition, examination, and analysis in order to assess the quality of the digital forensics in actual investigative reports. The framework can be used and further developed by LEAs as a template for improvement of digital forensics reporting.

2.2. Digital forensics investigation: Process and international standards

The reliability evaluation of the digital forensics process is the focus of this case study. The four sub-stages: Acquisition, examination, analysis, and reporting are considered the core of the digital forensics investigation. Consequently, only these essential stages were examined in this study. The scope of the study was also limited to three device types - computers, digital storage media, and mobile phones - for practical reasons.

2.2.1. Acquisition

Acquisition is the process of creating a copy of data from a particular data source (ISO/IEC 27037:2012). It is a vital step in the digital forensics process because it enables further examination while minimising the risk of data loss and alteration. Although there are a variety of different acquisition methods and tools, they all share the same essential steps: Read data from the source, write data to a target storage device, and verify that the data written are identical to the data read. The output of a digital forensics acquisition is usually a forensic copy and a hash value calculated from the original data.

A detailed description of the different acquisition methods falls outside the scope of this paper. However, it is worth noting that while manual extraction is a recognized method to extract information from a digital device (Ayers et al., 2014), this acquisition method differ from other acquisition methods in one key aspect: It does not actually capture the digital data stored on the device. Only the representation of data as provided by the device itself is captured. Manual acquisition also involves the probable modification, deletion, and overwriting of data since the person performing the acquisition must manipulate the device to access the data.

The assessment of the reliability of acquisitions in this study was based on three international guidelines and best practices: Interpol Global guidelines for digital forensics laboratories (Global guidelines, 2019, 5.1), ISO/IEC 27037 Guidelines for identification, collection, acquisition and preservation of digital evidence (ISO/IEC 27037:2012) and NIST Guidelines on Mobile Device Forensics (Ayers et al., 2014, 3.1, 4.2 and A.2.1). These standards were selected as they document fundamental digital forensics principles as recognized by the digital forensics community. While the NIST guide and ISO standard map general, technical and organizational conditions, the Interpol guidelines are focused on acquisition requirements of procedure specifically for law enforcement and define guidance notes for documentation of digital investigation. The standards were used to derive concrete reliability criteria and documentation requirements as described in section 3.3.1.

2.2.2. Examination

Examination involves converting raw data into a human readable form, structuring the data, and identifying potential digital evidence (Kohn et al., 2013). There are many methods and tools available to perform examination of data from computers, storage devices, and mobile phones (Global guidelines, 2019). Examples of common methods include searching (e.g. keyword searches, index searches, and string searches), data reduction (e.g hash analysis and data deduplication), and data recovery (e.g recovery of deleted files, decompression, and decryption). As a general rule, examination should be performed on a forensic copy (Global guidelines, 2019, 5.2.1).

The evaluation of the examination documentation in this study was also based on Interpol Global guidelines for digital forensics laboratories (Global guidelines, 2019, 5.1), ISO/IEC 27037 Guidelines for identification, collection, acquisition and preservation of digital evidence (ISO/IEC 27037:2012) and NIST Guidelines on Mobile Device Forensics (Ayers et al., 2014, 3.1, 4.2 and A.2.1). The specific reliability validation standard developed by European Network of Forensic Science Institutes (ENFSI) Best Practice Manual for forensic examination of digital technology was used to derive reliability validation criteria (ENFSI, 2015), and further elaborated to establish the minimum documentation requirements as detailed in section 3.3.2.

2.2.3. Analysis

Analysis relates to the identification and evaluation of digital evidence from sources of potential digital evidence (ISO/IEC 27042:2015). ISO/IEC 27042 defines digital evidence as 'information or data, stored, or transmitted in binary form which has been determined, through the process of analysis, to be relevant to the investigation' (ISO/IEC 27042:2015, 3.15). This means that potential digital evidence or digital artefacts cannot be considered digital evidence without, or prior to, analysis. Further, during analysis, digital artefacts and events are investigated and tested against hypotheses formulated in order to establish the probative strength of the digital evidence (ENFSI, 2015, para 11). The OSAC group classifies three types of data analysis: Organize observed traces to disclose the most likely operational conditions or capabilities (*functional analysis*), patterns in time (*temporal analysis*), and linkages between entities - people, places, objects - (*relational analysis*) (Pollitt et al., 2018). There are various content-dependent methods, tools, and test setups that can be used for the different analysis types.

The assessment of reliability criteria for analysis in this study was based on NIST Guidelines on Mobile Device Forensics (Ayers et al., 2014, 3.1, 4.2 and A.2.1), Interpol Global guidelines for digital forensics laboratories (Global guidelines, 2019, 5.1), ISO Guidance on assuring suitability and adequacy of incident investigative method (ISO/IEC 27041:2015), ISO Guidelines for the analysis and interpretation of digital evidence (ISO/IEC 27042:2015) and the ENFSI Best Practice Manual for forensic examination of digital technology (ENFSI, 2015). These standards were used to define the minimum reliability validation documentation of the analysis stage as described in section 3.3.3.

2.2.4. Reporting

Reporting and documentation is an integral part of all forensic work and the way digital evidence is communicated. Details about all actions performed, all considerations made, and all results obtained during digital forensics investigation should be documented (ISO/IEC 27037:2012; Casey, 2018; Ayers et al., 2014; CoE, 2014). Consequently, this study expected reports of digital forensics actions to be accurate and to provide comprehensive accounts of the digital forensics investigation performed during a criminal investigation, i.e. that acquisition, examination, and analysis were documented at minimum to demonstrate compliance with international standards as described above. The complete account of a digital investigation might be spread across multiple documents. Some devices might have been acquired, examined or analysed multiple times with different methods and tools, and by different people. However, forensic reports were expected to link the various stages and actions together as a logical sequence for other parties in the criminal proceedings to be able to follow and understand what has been done, and to trace the digital evidence back to its origin. Hence, the reports should collectively contain the minimum documentation for reliability validation of the acquisition, examination, and analysis as described here.

2.3. Reliability validation enabling framework (RVEF) for digital forensics investigation

Reliability is a property of the digital forensics process related to a consistent intended behavior and results. It equals reproducibility and validation of the method (Risinger, 2018). A test that produces the same results on successive applications is said to be reliable (Gross and Mnookin, 2003). In previous work a *Reliability validation enabling framework* (RVEF) was created. The RVEF defines four validation criteria (Stoykova, 2021a, p.12), *data set*, *tool*, *method*, and *examiner work* that must be documented at technology, method, and application level to create audit trails of digital forensics actions (Stoykova, 2021b, p.81) Such minimum documentation enables validation of routine and easily verifiable tasks in

acquisition, as well as robust lab testing during evidence examination and analysis. The RVEF provides a background for the method employed in this study. As the standards and guidelines in section 2.2 are extensive and describe validation procedures, RVEF defines the minimum requirements at technology, method, and application level that can satisfy a validation procedure at any stage of the digital forensics process. It is expected that the law enforcement reports contain such minimum reliability information.

2.3.1. Technology level

The technology level documentation must provide information about the employed functionality of the automated setup. It is recommended to use only validated tools which should be re-validated with each software update (Global guidelines, 2019, 3.4) (Ayers et al., 2014, 3.4) At minimum the documentation of commercial software, in-house tools, and scripts must include the name, version, configuration, and functions used (ENFSI, 2015, 4.2 and 6.6). References to previous validation and verification testing, and stating known errors e.g. data interpretation limitations, are also important, along with known bugs, and the tool's ability to report errors (ENFSI, 2015, p.23).

2.3.2. Method level

A validation of the method is an 'assessment of whether a standardized sequence of steps, often employing digital forensic tools, leads to a reliable result' (Hughes and Karabiyik, 2020, p.5). There is an overlap between the validation requirements at technology and method level as a tool can encompass several methods (functions) or vice versa (Marshall and Paige, 2018). Therefore, the documentation should include the tool function used, in order to derive possible algorithm and software implementation errors or method limitations (ENFSI, 2015). From the documentation it should be possible to determine if an appropriate scientific method, technique or procedure was followed (ISO/IEC 27037:2012), if the method meets the requirements of the investigation and if it has been appropriately tested (ISO/IEC 27042:2015). The examiner might refer to peer reviewed methods, established practices, or previous work (ENFSI, 2015, 4.3). Consequently, the minimum information necessary for reliability validation includes pre-processing for input, algorithm, and feature selection method as well as known limitations.

2.3.3. Application level

At application level, the examiner must ensure that the methods and tools work correctly and as intended. The forensic task defines the scope of automated processing (Global guidelines, 2019, 5.2.3.3) and influences the selection of tools and methods (ENFSI, 2015). Further, the methods used for each forensic task should be validated (ISO/IEC 27041:2015, 5.5.2) (ISO/IEC 27042:2015, p.7). Descriptions of the data set, including file path and format, integrity verification (hash) function (Global guidelines, 2019, 5.1.2.3 and 5.1.3.3) (Ayers et al., 2014), and specification of the secure storage (European Network of Foren, 2015, p.12) are all preconditions for validation. To validate the examiner work at the application level, the documentation must contain at minimum a description of subjective measurements e.g. hypotheses, assumptions, and decision based on expert knowledge (ISO/IEC 27042:2015, 6.4) (Global guidelines, 2019). Examiner interaction with the tools must be traceable and include justification for the selection of tools (ISO/IEC 27041:2015, 5.11.1) and methods (ISO/IEC 27041:2015, 5.6.2 and 5.6.3)(Ayers et al., 2014), and information about algorithms and the feature selection process according to the task, data set characteristics, and methods limitations. Concluding remarks must express confidence level, precision, and accuracy, and clearly separate facts from inferences (ENFSI, 2015, 13.4) (ISO/IEC 27042:2015, 9.2), as well as interpretation of results (ENFSI, 2015, para 12)

In order to simplify the examined requirements, the RVEF documentation schema is further detailed in section 3 where the minimum evaluation requirements for each of the reliability criteria are described.

3. Method

The method in this study describes the minimum information that must be documented in order to enable reliability validation of any processing stage in a digital forensics investigation. The reliability validation criteria and evaluation requirements were used to evaluate the data set and assess the reliability of the current digital forensics work in randomly selected cases from the Norwegian police.

The validity of this study rests on the information documented and presented in the reports. It is possible that the reports do not accurately or comprehensively describe the digital forensics examinations performed. However, the reports were all used to inform both the prosecution and the defense prior to trial, and might have been presented as evidence in court. Consequently, any errors and deficiencies in the reports might have impacted on the reliability of the evidence and therefore on the quality of the criminal investigation and procedure. They might also have affected the opportunity of the parties to examine and challenge evidence on valid grounds. Therefore, the reports were considered representative and valid sources for this study.

3.1. Hypothesises

Hypothesis 1. Each stage of the digital forensics process (acquisition, examination, analysis and reporting) was performed according to international digital forensics standards, see section 2.2. At minimum, the documentation reports on technology, methodology, and application according to the reliability validation matrix, see section 2.3.

Hypothesis 2. The digital investigation actions, and their results, were documented such that the origin, integrity, and interpretation of the digital evidence can be cross-examined.

3.2. Data collection and sampling

The study was based on reports collected from a random set of criminal investigations in Norway. The set was constructed from cases where the incident under investigation took place between January 1st, 2016 and December 31st, 2019. A new penal code came into force in Norway on October 1st,2015.[1] This led to changes in the way cases were categorised and in the way crime statistics were calculated. With the last months of 2015 essentially becoming a transition period (Stene, 2017), cases were selected from the time after the new code and statistics had settled.

The case population was constructed from homicide (section 275 NPC) and sexual assault (sections 291 - 293 NPC) cases. These were selected partly because of the nature and severity of the

[1] The Norwegian Penal Code (NPC) of 2005, English translation available at https://lovdata.no/NLE/lov/2005-05-20-28.

Table 1
Case metadata extracted from the criminal records database and used in sampling.

Data field	Description
Case number	Identifying serial number unique to each case.
Incident start & end date	The dates when the incident took place.
Registration date	The date the case was registered.
Current phase	The current phase of the case: Closed, investigation ongoing, processing by the prosecution authority or by administration.
Statistical code	Categorisation code based on type of crime.
Processing district	Police district or special agency processing the case.
Partial case	Indicates if the case is split into multiple parts.
Main case	The case number of the main case. (Only attached/subordinate cases have a main case number.)
Attachment case	Indicates if the case is attached (subordinate) to another case.
Attached cases	The number of attached (subordinate) cases.
Lawful decision	Code categorising the end result of the case.
Date of lawful decision	The date of lawful (final) decision.
Prosecution decision	Code categorising the prosecution authority recommendation.
Clearance code	Code categorising the end result of the case related to one of the defendants.

crimes they represent,[2] and partly because digital evidence was assumed to be relevant to the investigation of these cases.[3] These crimes are among the most serious in society and are prioritized for investigation by the Norwegian Director of Public Prosecution (Riksadvokaten, 2020).Further, digital evidence has been shown to be relevant in previous cases of this nature. According to The UK National Police Chiefs Council, 90% of the criminal investigations have digital element (The UK National Police Chiefs Council, 2020).

Note that the absence of digital forensics reports does not mean that no digital forensics investigation was conducted; only that no report was made available in the case management system. It is reasonable to expect that the existence of a report means that some work was performed. This study examines the quality of that work as presented in the reports.

The target sample size was 21 cases and the following selection process was employed: First, lists of cases including case metadata (see Table 1) were extracted from the criminal records database.The metadata was used to pre-screen and compile the population such that only cases from the relevant period (incident and registration dates) and of the relevant categories (statistical code), in which the investigation led to an indictment (decision), and which was subsequently tried in court leading to a verdict (lawful decision and clearance codes) were selected. Further, only independent cases, i.e cases which were either not joined with other cases (attachment case), or which were designated as the main case in a complex of multiple cases (main case and attached cases), were included. Using Microsoft Excel, a table was created for each of the relevant case categories. With each row containing the metadata from one case, a random number was assigned using the RAND ()-function. The table was then sorted by this number and cases were sampled from the top of the list until the target sample size was reached. Once sampled, information about all the seized items[4] registered with the case was extracted from the case management system. This information was gathered from search and seizure reports—a special kind of report generated when information about seized items are registered with the system. If at least one computer, mobile phone, or digital storage device was registered, the case was accepted. After sampling 25 cases, we reached a total of 21 from 11 different police districts.[5] Where the incident took place or which police district investigated the case was not taken into account. Then, for each selected case, all reports concerning computers, mobile phones, and storage devices were extracted from the case management system. All the extracted reports were pseudonymized according to a data protection impact assessment (DPIA) conducted prior to initiating this study[6] to ensure that the internal structure and logic of the reports were preserved without exposing personal data.

The reports were categorised into five different categories: *Acquisition reports* describe the forensic acquisition of digital data. *Examination reports* describe the processing prior to, and the search for data of relevance to the case, i.e. the search for possible digital evidence. *Analysis reports* document the identification and evaluation of digital evidence. *Content reports* contain digital data only, or reproductions or representations of such, e.g. text, tables or pictures, found during examination. *Photography reports* consist of photographs of digital data, e.g. photographs of a computer or mobile phone screen. The latter were not considered a digital forensics report and it was not a category initially included in the design of the study. Rather, it was added during data collection after multiple such reports were encountered and it became necessary to report this finding in a consistent manner.

3.3. Reliability validation method

The reliability validation methodology was divided according to the minimum digital forensics process, i.e. acquisition, examination, and analysis. Reporting is usually considered a distinct process step. However, since the reports were the object of the study only the three core stages were considered.

The assessment did not follow the logic of the reports. Because the information necessary to conduct the study was scattered across multiple reports, the evaluation was performed in two steps. First, the relevant information was collected from all the reports concerning digital evidence in each case. Subsequently, qualitative assessments of the core digital forensics process steps (acquisition, examination, and analysis) were made using the evaluation

[2] Homicide, sexual assault and aggravated sexual assault are all subject to a penalty of imprisonment of up to 21 years.

[3] Initially, cases concerning the illicit handling of authentication details, computer programs, etc (section 201 NPC), intrusion into computer systems (section 204 NPC), and violation of the right to private communication (section 205 NPC) were also included in the study. However, during sampling no digital items were found to have been seized in any of these cases in the population.

[4] According to Sec. 203 Norwegian Criminal Procedures Act (NCPA).

[5] The Norwegian police is divided into 12 police districts. Criminal cases are primarily handled by the police district in which the incident took place. Some special cases are investigated by the National Crime Investigative Service (NCIS) or the National Authority for Investigation and Prosecution of Economic and Environmental Crime (ØKOKRIM). No such special cases were included in this study. Cases under the jurisdiction of the Police Security Services (PST) were not included, either in the population or in the sample.

[6] The DPIA was performed with the assistance of the Norwegian center for research data (NSD), reviewed by NTNU privacy officer and accepted by the Head of the Department of Information Security and Communication Technology at NTNU.

Table 2
Reliability criteria for acquisition.

Reliability criteria	Evaluation requirement
Mandate	Justification for acquiring data, e.g. investigative or digital forensics hypothesis
Data source description	Device manufacturer and model, unique identifiers (e.g. serial number, IMEI), inventory identification numbers and acquisition space (i.e. input data)
Tool description	Name, version and device specification (where applicable) of all tools (software and hardware) used during acquisition
Method description	Tool and function specification, file format, and integrity verification (e.g. hash algorithm and value)
Examiner	Time of acquisition, Examiner identification (e.g. name, username or similar), action logs and justification for selected tool and method
Acquisition results	Acquisition output, tool and method output and storage location

Table 3
Reliability criteria for examination.

Reliability criteria	Evaluation requirement
Mandate	Scope of the task based on investigative hypothesis for discovery of relevant data
Data set description	Description of the data to be examined including storage location, path or link to acquisition output
Tool description	Name, version and device specification (where applicable) of all tools (software and hardware) used for examination
Method description	Method type and specification (e.g. Function, algorithm, etc.), examination space (i.e. the area of the data set to be examined) and Parameters (e.g. Keywords, encoding, magic numbers, known hash values, etc.)
Examiner	Time of examination, Examiner identification (e.g. name, username or similar), action logs and justification for choice of tool and method according to data set characteristics and forensic task
Examination results	Findings (i.e. potential digital evidence, origin/forensic path and encoding/interpretation), inferences and assessment of the findings, and examination conclusions relative to the mandate

Table 4
Reliability criteria for analysis.

Reliability criteria	Evaluation requirement
Mandate	Digital forensics hypothesis and type of forensic analysis to be performed (functional, temporal or relational) (Pollitt et al., 2018)
Data source description	Specification of the data to be analysed (i.e. storage location, path or link to acquisition and examination output) and contextual information (e.g. incident information, system information or nature of the source of the digital artifact)
Tool description	Name, version and device specification (where applicable) of all tools (software and hardware) used for analysis
Method description	Type of digital forensics analysis performed (e.g. file systems analysis, network forensics analysis, etc.), analysis space (i.e. the area of the data set analysed) and description and specification of the analytical process
Examiner	Examiner identification (e.g. name, username or similar), action logs and justification for choice of tool and method according to analysis type and process
Analysis results	Digital evidence (data and origin/forensic path), interpretation of the evidence, inferences and assessment of the evidence and analytical conclusions relative to the mandate

requirements specified in Tables 2–4. As the evaluation requirements are the minimum necessary information for reliability assessment, the absence or partial documentation of any of the reliability criteria means that the process step can not be validated or reproduced. From the evaluation, each reliability criteria was assigned one of the following values depending on the data in the reports:

- Yes: If all evaluation requirements were met
- Partial: If at least one (but not all) of the reliability criteria were documented
- No: If none of the reliability criteria were met or no data was presented

3.3.1. Reliability criteria for acquisition

The selected criteria that needs to be documented at minimum to evaluate the reliability of the acquisition stage are: *mandate*, *data source*, *tool*, and *method description*, *examiner* and *acquisition results*. A mandate provides the boundaries for actions, including the authority and the reasons and motivation for their execution. The former ensures compliance with criminal procedure, while the latter hinges on the hypotheses they are set to inform (Andersen et al., 2021). Due to the scope of this study, only the requirements in respect to the hypotheses were considered. Authorisation requirements and legal procedures were assumed to have been followed.

For data acquisition a mandate is required to determine and justify how to proceed. This study looked for statements which informed the reason for acquiring data from a particular device, including the scope of the acquisition. Such statements could be derived from an investigative hypothesis, a digital forensics hypothesis, or a specific task (e.g. to look for a particular type of data). For the purpose of this study, an investigative hypothesis was understood to be concerned with locating digital data (digital artifacts) suitable to inform the incident under investigation, and a digital forensics hypothesis was understood to be part of the scientific methodology which aims to determine the probative strength of the digital artifacts themselves.

Devices from which data is acquired must be sufficiently described in order to trace the data to its source. The acquisition space should be documented as different acquisition methods can acquire different parts of the device data (e.g. partitions, volumes, allocated or unallocated space). In this study, the presence of a device make and model description was considered insufficient to uniquely identify a device. At least one unique device identifier, e.g. a serial number or an IMEI,[7] along with the inventory number

[7] International Mobile Equipment Identity.

assigned during legal seizure was required (Ayers et al., 2014, 5.1). This minimum documentation enables the tracking and linking of a device from seizure, e.g. at the crime scene or from a person, to acquisition and beyond.

Digital forensics tools - both hardware and software - support different device types and models, and they have different functional limitations. The tool description informs if the tool is suitable for the task it was applied to. For this study, the name, version and device specification were allconsidered necessary to provide sufficient information for tool suitability validation and to enable performance assessment.

Digital forensics practitioners rely heavily on software. And as Horsman details, tools can be a major source of error and uncertainty leaving a great deal of the quality of forensic work in the hands of software developers. Being aware of and distinguishing clearly between tool errors, user errors and tool limitation is therefore key to determine the reliability and validity of forensic work (Horsman, 2018b). Without detailed information on what tools were used and how, it is impossible to assess these properties for any of the steps in the digital forensics process.

The acquisition method overlaps with the tool as it can be embedded as a function of the tool or encompass functions of multiple tools. Specification of the function, file format and integrity verification are the minimum characteristics necessary to validate the method used for acquisition. Reporting the integrity verification method and value are crucial to enable acquisition validation and to ensure that the examinations and analyses are performed on a true copy of the original data. All these criteria were required in the study for the acquisition method to be considered sufficiently documented.

Examiner criteria refers to when, by what actions and by whom an acquisition was performed. Reporting the date and time of acquisition enables the evaluation of what technology and methodology was available at the time of acquisition, and impacts the justification for using it. The identification of the examiner provides information for validating the competency and the chain of custody. Again, all these criteria were required to considered the examiner to be sufficiently documented.

One of the important tasks of the examiner is to justify the selection of the acquisition method and tool. The choice depends on the device specifications, tool compatibility, an assessment of data availability (e.g. encryption) and whether there is a need to acquire volatile data.

The acquisition output is the forensic copy of the digital data acquired from the source device, while the tool and method output refer to the run-time log and error report produced by the tool or observed by the practitioner. Storage location refers to the secure storage device, location and path where the acquired data was stored.

3.3.2. Reliability criteria for examination

The examination mandate directs what to look for. Thus, it is motivated by and based on the investigative hypotheses or a digital forensics hypothesis (ISO/IEC 27042:2015, 8.3), 8.3]. The mandate impacts the selection of the methods and tools, the pre-processing and structuring of data for identification and classification of potential digital evidence, and selection of relevant data characteristics. This study looked for statements dictating a specific task and its scope e.g., specific data, data types, data formats, dates, devices or accounts.

The data set description informs what data is to be examined. It can include parts, a whole or several acquisition data sets. An accurate storage path or similar is necessary in order to identify what data was examined.

The requirements for examination tool description are the same as for acquisition described above.

Examination methods describe what and how different tools, functions, and algorithms were employed. Digital forensic tools—like all other software and hardware—can have bugs and errors, and the functions used might not be suitable for the task. This can cause tool errors and user errors respectively as described by Horsman (Horsman, 2018b). Therefore, the specification of the methods employed and of the examination space (the exact area of the data set on which a method was employed) is required. In addition, correct functioning algorithms might produce erroneous outcomes if the wrong parameters are set. Explicitly reporting the selected parameters (e.g. keywords, encoding, magic numbers, known hash values, etc.) is a minimum requirement to assess the method and validate the results.

Investigations evolve over time. Examinations can become irrelevant or they may have to be repeated as the investigation progress. Identifying the examiner, what actions were taken and when is essential for continuity, re-examination, and the assessment of relevancy, completeness and possible effects of cognitive bias of the examination. As already stated, the choice of tools and methods for examination depends on the data set characteristics and the forensic task. A justification of selected examination by the examiner ensures that what was done was proportional and did not exceed the scope of authorisation.

Examination results are the interpretations of the data by tool or examiner. This includes findings of potential digital evidence, origin/forensic path and encoding/interpretation, inferences and assessment of the findings, and examination conclusions relative to the mandate.

3.3.3. Reliability criteria for analysis

The evaluation requirements for the analysis stage of the data processing are similar to the examination. The difference is that before the mandate and data source for analysis are defined, the outputs from acquisition and examination has to be enriched with contextual information about the investigation in order to define the scope of the analysis.

4. Results

4.1. Reports

A total of 124 reports concerning digital devices and investigation of digital data were collected from the 21 sampled cases. 62 reports were from homicide cases and 62 were from sexual assault cases. Table 5 shows the distribution of the reports across the five report categories. Some of the reports fit into multiple categories, e.g by describing both acquisition and examination. Thus, the sum in Table 5 is higher than the number of documents extracted.

Table 5
Number of categorised reports collected and assessed per case type.

Case type	Acquisition	Examination	Analysis	Content	Photography	Sum
Homicide	24 (34%)	35 (50%)	2 (3%)	3 (4%)	6 (9%)	70
Sexual assault	7 (10%)	32 (48%)	1 (1%)	23 (34%)	4 (6%)	67
Total	31	67	3	26	10	137

Table 6
Number of devices identified per case type.

Case type	Computers	Mobile phones	Storage devices	Sum
Homicide	27 (23%)	71 (59%)	22 (18%)	120
Sexual assault	6 (9%)	45 (67%)	16 (24%)	67
Total	33 (18%)	116 (62%)	38 (20%)	187

Table 7
Type of documentation found about the identified devices (n = 187).

Acquisition report	Examination report	Analysis report	Number of devices
			71 (38%)
•			32 (17%)
	•		40 (21%)
		•	0 (0%)
•	•		41 (22%)
	•	•	1 (1%)
•		•	1 (1%)
•	•	•	1 (1%)

4.2. Devices

A total of 187 digital devices were identified in the 21 cases. 181 of them were documented in search and seizure reports, while a further 6 devices were identified during the review of the reports. Table 6 shows the distribution of device types per case category, while Table 7 shows a summary of what we found to have been reported about the devices.

The results show that from the three device types studied, mobile phones was the dominant device type in both homicide (59%) and sexual assault (67%) investigations.

4.3. Acquisition reliability assessment

Information about acquisition was available for 75 out of the 187 devices (40%). The information was scattered across reports describing either single device acquisition, acquisition of multiple devices, or reports concern both acquisition and examination. When assessing the acquisition reliability, information regarding each device was complied and assessed from all these reports. The photography reports were excluded. In one of the cases, an acquisition report described how one device was physically damaged such that acquisition was impossible. Tool and method description was considered *not applicable* in this instance.

31 of the total 124 reports (25%) were considered acquisition reports and only 14 out of 21 cases (67%) had such reports. In addition, 10 reports included photographs of the device screen or screenshots. In these instances, the examiner was assumed to have browsed the device manually (LIVE examination) in order to access and display the captured data. Forensic acquisition reports were only found concerning two devices from those reports. 15 acquisition reports concerned more than one device. In some of these it was not possible to discern which tool or method was employed for each device. No acquisition report was found for 106 (57%) devices.

None of the acquisition reports provided sufficient information to validate the reliability of the process step in question.

Table 8 shows the breakdown of the acquisition reliability assessment.

4.3.1. Mandate

An investigative mandate was specified for 13 reported acquisitions. These focused on communication between the suspect and the victim, digital events around the time of the incident and media files (e.g. pictures and videos) concerning the incident. For 50 reported acquisitions, no scope or justification for acquisition was specified. No mandate was specified for the remaining 11 reported acquisitions. None of the mandates specified a digital forensic hypothesis for the acquisition.

Table 8
Number of reported acquisitions per acquisition reliability criteria (percentage calculated per criteria).

Reliability criteria	Yes	%	Partial	%	No	%	N/A	%
Mandate	13	18%	50	68%	11	15%	0	0%
Data source description	4	5%	70	95%	0	0%	0	0%
Tool description	17	23%	38	51%	18	24%	1	1%
Method description	0	0%	55	74%	18	24%	1	1%
Examiner	1	1%	73	99%	0	0%	0	0%
Acquisition results	2	3%	51	69%	21	28%	0	0%

4.3.2. Data source description

When an item is seized it must be registered with the case management system. The registration includes a description of the item (e.g. manufacturer, model, colour, etc.), any serial numbers or other unique identifiers, information about where the item was found and where the item is stored. As part of this registration, the seized item is assigned a unique inventory number. A review of the seizure reports showed that not all the information were included in the seizure reports[8] and that the description of the devices was scattered across several different reports. The study also found that not all reports concerning acquisition referred to the assigned inventory number. Rather, a device description (e.g. manufacturer or model name), the ownership of the device (e.g. the suspect's computer) or an IMEI or serial number was often used in place of the inventory number.

Tracing the information about the device across all case documentation, the study found that only four devices were documented according to the requirements. The remaining 70 (37%) devices all fulfilled at least one of the requirements. Three reported acquisitions referred to an inventory number but were acquired without any information about the make, model or unique identifiers. 42 reported acquisitions did not include a unique device identification.

4.3.3. Tool description

The study found that both commercial and open source tools were used for data acquisition. 17 of the reported acquisitions specified both the tool name and version. One report described the acquisition of 16 different devices using three different tools, but did not specify which tool was used on which device. A further 22 reported acquisitions provided only a partial description of the tool used (i.e. partial tool name, missing version number, etc.). Another eight reported acquisitions described the use of 'special equipment' to acquire data. No further information was provided about this equipment. Only seven reported acquisitions mentioned the use of a physical write blocker, and 19 of the reported acquisitions provided no information regarding the acquisition tool.

4.3.4. Method description

For 52 of the reported acquisitions, the function used to acquire data was not specified. Seven of the reported acquisitions provided sufficient information to accurately determine the function employed. The remaining 14 reported acquisitions referred to imaging, physical or logical acquisition or similar, but provided no further information to determine the exact functionality of the tool used.

[8] A thorough review of the seizure reports was outside the scope of this study.

Further, for 50 of the reported acquisitions, no mention was made on how or if the integrity of the acquisition was verified. 18 of the reported acquisitions stated that MD5 was used, but presented no actual hash value. For three reported acquisitions both MD5 and SHA1 were reported, but without a value. Only for two reported acquisitions were the use of MD5 specified and the hash value provided.

Note that the functions available to acquire data from a device depends on the type of device and the tool employed. The assessment of the function used to acquire data did take into consideration whether the data source and tool description allows for an accurate determination of what function was used during acquisition. For example, in one case both the device make and model, and the tool name and version was specified, along with the statement that the function 'advanced logical acquisition' was used. In these instances the reliability criteria for method was consider fulfilled. However, in a different case, the tool specification was incomplete. Thus, even though the function was described, it could not be deduced which method was actually used to perform the acquisition and the description was considered insufficient.

4.3.5. Examiner

All the acquisition reports specified the name and rank of the author. Two acquisition reports specified data and time of acquisition, and 17 only the date. For 55 devices there was no information about when the acquisition was performed. Out of the 74 reported acquisitions, 39 did not describe any of the actions performed while 35 gave a partial description of some of the actions. 65 reported acquisitions did not specify any justification for the selected method or tool.

4.3.6. Acquisition results

38 reported acquisitions did not describe the acquisition output or the file format used. 28 acquisitions referred to mirror image files and two refer to acquisition files. Another two named .BIN-files as the resulting output, but insufficient information was provided about the tool or method used to allow inferences about the acquisition output. One of the reported acquisitions specified that the output was a report generated by the acquisition tool.

None of the reported acquisitions provided detailed information about the tool and method output, e.g. run-time logs or tool results report. Three of the reported acquisitions specified that no errors were reported during acquisition, while one noted that an error occurred and that the practitioner was unable to connect to the device. No details were provided about the error.

Further, none of the reported acquisitions provided storage location information. However, 22 of the reported acquisitions contained comments on how or where the acquired data was stored, e.g. saved to external USB drive, saved to police hard drive, or stored at organisational unit. These were considered to partially satisfy the reliability criteria.

Finally, one reported acquisition was cancelled due to an error while connecting to the device and one was cancelled due to a missing PIN-code.

4.4. Examination reliability assessment

17 out of 21 cases (81%) contained at least one examination report. Of the total 124 reports, 67 (54%) were categorised as examination reports. Since some reports concerned more than one device, the number of reported examinations (104) is greater than the number of examination reports. Also, while 80 of the 187 devices (43%) were examined at least once, some devices were examined more than once leaving 107 devices (57%) not reported as examined. Finally, 53 reported examinations found potential relevant evidence data.

Table 9
Number of reported examinations per examination reliability criteria (percentage calculated per criteria).

Reliability criteria	Yes	%	Partial	%	No	%	N/A	%
Mandate	63	60%	8	8%	33	31%	1	1%
Data source description	0	0%	38	36%	64	61%	3	3%
Tool description	12	11%	38	36%	54	51%	1	1%
Method description	0	0%	29	28%	74	70%	2	2%
Examiner	0	0%	104	99%	0	0%	1	1%
Examination results	25	24%	78	74%	0	0%	2	2%

None of the examination reports provided sufficient information to validate the reliability of the process step, as most evaluation requirements were only partly met or not present in the reports. Table 9 shows the breakdown of the examination reliability assessment.

4.4.1. Mandate

33 reported examinations were conducted without specifying any mandate. Eight partial mandates included general statements such as 'examine acquisition file' and 'find relevant data' without further scope or task specification. In 63 reported examinations, the task and scope was stated or could be directly inferred as relevant to the case. For example, the examiner specified concrete data types, persons and objects of interest, time period, or communication between named persons.

In one case, the examinations of one computer were described as inadequate during trial at district court. New examinations ordered prior to trial at the appeals court uncovered relevant data. The mandates from defence lawyers and district attorney formulated for this second round in court appeared more specific than those constructed during the initial investigation in respect to the forensic task.

4.4.2. Data source description

None of the 104 reported examinations provided a sufficient description of the examined data source. This was primarily due to missing information about the storage location. Five examinations, all related to the same case, referred to data stored on external CDs and DVDs. Only one provided information about the storage location and path. Another examination specified the filename of the examined data and one referred to the forensic path of the examined data - neither providing information about where the file nor forensic image was stored. One examination described that it was performed on a mirror image copy, specifically the user area of a single user on a file system where a particular operating system was found, but gave no path or storage device information. A further 27 reported examinations referred to acquisition or image files, or a type of data (e.g. communication events and videos) without specifying file names, formats or paths. Eight examinations were reported as manual examinations, i.e. LIVE forensics.

4.4.3. Tool description

12 reported examinations specify the tool name and version, and therefore marked as sufficient. However, none of them specified what devices (e.g. which computer or write-blocker) were used. In several reports, including reports concerning multiple devices, multiple tools were listed without further information on why they were used or on which data set they were employed. 54 examinations did not provide any data about the tools used, while 32 provided only the name of the tool. Four examinations referred only to the tool provider.

4.4.4. Method description

None of the reported examinations gave sufficient information on the method used to enable the process to be repeated. In 16 examinations, the examination space was described as application data (e.g. WhatsApp or Skype), call logs or communication data without further specification of the data location or forensic path in the data set. 13 examinations mentioned searching as the method, but neither the type of search, nor the search space or the search parameters were reported. Two examinations specified keyword search but the keywords were not provided.

The eight manual examinations did not specify how the examinations were conducted.

74 reported examinations provided no information about the method used.

4.4.5. Examiner

All the examination reports included the name and rank of the author. We accept this as identification of the examiner. However, out of the 104 reported examinations, 90 did not describe any of the actions performed while 14 gave a partial description of some of the actions. 99 reported examinations did not provide a justification for the method or tools used. Only two examinations gave such justification. Eight examinations gave the date of the examination, four gave both a start date and an end date, and one specified both date and time.

4.4.6. Examination results

All the examinations provided some information about their results. 25 concluded that nothing of relevance was found, or that no data was available. These were the only ones found to sufficiently document their results. A further 10 examinations concluded that nothing of relevance was found, yet findings (i.e. content data from the examined device) were still described.

Only in one examination which identified relevant data was the forensic path to the find provided. 33 examinations described or reproduced relevant content data, but provided no forensic path or reference to its origin. Six examinations described or reproduced content data, but failed to provide information on if or how the data was relevant or not.

Findings were primarily provided through descriptions or partial reproduction of the content data (e.g. table and pictures pasted into the report), or screenshots of content data. Further, no conclusion was found in 26 examinations due to missing mandate. One examination concluded that the data on the device was irrelevant based on the last change date/time of the device. The date and time were presented, but no forensic path or information about the origin of the data was provided.

12 cases contained content reports (see Table 5). These reports were either extraction reports generated by a fo rensic tool, a report produced by a digital forensics examiner consisting primarily of content data, or illustration reports consisting of photographs or screenshots (reported in Table 5 as photography reports). Only in two cases did the examination reports or the content report refer to each other such that the link between the examination and the content data was maintained.

4.5. Analysis reliability assessment

53 reported examinations yielded potential digital evidence, i.e. provided data without stating explicitly that nothing of relevance was found. The study expected to find analysis reports which establish the relevance, meaning and strength of the potential evidence identified of all these examinations. Yet only three such reports in two different cases were identified. Table 10 shows the breakdown of the analysis reliability assessment.

Table 10
Number of reported analyses per analysis reliability criteria (percentage calculated per criteria).

Reliability criteria	Yes	%	Partial	%	No	%
Mandate	1	33%	0	0%	2	66%
Data source description	0	0%	2	67%	1	33%
Tool description	0	0%	1	33%	2	67%
Method description	0	0%	2	67%	1	33%
Examiner	0	0%	3	100%	0	0%
Analysis results	0	0%	2	66%	1	33%

None of the analysis reports provided sufficient information for a reliability assessment. Only one provided a statement of purpose and a partial description of the tool used. Two analysis reports mentioned the method used, but none gave information about where the data was stored, version of the tools used, time or date when the analysis was performed or justification for the method or tool used. Two of them gave a partial description of some of the actions performed and described the findings, but no actual data was provided. Nor were there any information on data interpretation.

4.6. Observations

Information about each item and its acquisition, examination, and analysis were scattered across multiple reports. The source, origin, or authenticity of the digital artefacts were not always reported. References to devices were not consistent either in their existence or format.

The reports were produced in isolation and lacked consistency across the different digital forensics process steps. The absence of reference to previous stages or processing steps made it hard to establish the digital chain of custody and chain of evidence. There were no audit trails that could be used to trace the processing steps backwards for a particular piece of data found to be of relevance.

Multiple manual examinations (live forensics) were performed with no justification or detailed description. In one instance, the report stated explicitly that 'no changes were made to the mobile phone', but the statement was not supported by any documentation or explanation.

Contradictions and copy-paste errors were found. For example, in one instance the identifying number in the report title did not match the number used in the rest of the report. In another case, a separate report described how two mobile phones were mixed up due to a copy-paste error. In yet another case, an acquisition report stated that data acquired from a device was found to be encrypted. An examination report described data found on the same device without any mention of if, and how, the acquired data was decrypted or if it was accessed by other means.

It was noticeable that no digital forensics terminology for the core forensic processes was used consistently. Inaccurate terminology such as 'draining data',[9] understood to mean acquire data, and 'peruse data'[10] understood to mean examination, were used in several reports. The term 'mirror image'[11] was also used in at least 23 reports about acquisition. This term is obsolete and imprecise. It used to refer to a bit-by-bit copy of all data on a storage medium. However, obtaining such a copy from today's storage devices is practically impossible (Nikkel, 2016; Kumar, 2021). For example, over-provisioning and garbage collection implemented in firmware

[9] Our translations from Norwegian *tømme/tappe data*.
[10] Norwegian *gjennomgang*.
[11] Norwegian *speilfil*.

on solid state drives effectively hides data from all but those with detailed information on the implementation of these functions. Similar issues apply to most if not all types of hand-held devices.

5. Discussion and recommendations

It was expected (Hypothesis 1) that each stage of the digital forensics process was performed according to the digital forensics standards (see section 2.2), and that a minimum of documentation was provided on technology, methodology, and application according to the reliability validation matrix (see section 2.3). It was further expected (Hypothesis 2) that the digital investigation actions, and their results, were documented such that the origin, integrity, and interpretation of the digital evidence could be cross-examined (see section 3.1). The results show clearly that the reported digital forensics actions were not performed according to international digital forensics standards, nor did the reports contain the minimum documentation at technology, method, and application level to enable reliability validation. The results also show that none of the cases were sufficiently documented to enable the assessment of reliability of the digital evidence, to trace the digital forensics actions performed on each item, or to link the digital evidence to its respective origin and source. Further, none of the cases were shown to comply with digital forensics methodology, justify the methods and tools used, or validate tool results and error rates.

From the results and observations it appears to be little or no consistency in the reporting and no forensic process or methodology was documented. This lack of consistency between reports related to the same device or data set makes it nearly impossible to reliably associate data artifacts with their respective data source. Hence, the reliability of the digital evidence cannot be determined.

The use of screenshots and photographs to acquire and preserve data seemed prevalent. While this might be a useful way to present data, pictures of data is not a forensically sound method of acquisition. It might be the only option in emergency situations, however, digital forensics conclusions cannot be drawn from such data. Hence, this practice should not be used routinely.

In some acquisition reports, the only information about the method and tools used was phrased as 'special equipment'. With no further information it is impossible to assess and review the results of these acquisitions. The reason for omitting such information was unclear. The lack of such documentation has a profound negative impact on all subsequent forensic actions and on the possibility to establish the reliability of, or to challenge the digital evidence on reasonable grounds.

Investigative and forensics actions were performed in parallel without a clear distinction between the two. For example, digital forensics personnel was requested to provide information rather than to examine or analyse investigative or forensic hypotheses. In some cases data was prepared by digital forensics personnel for later examination by an investigator. However, no further digital forensics examination or analysis was reported. Consequently, digital forensics practitioners were performing single tasks to supply the investigators with data or information to advance investigative objectives rather than to establish relevance and probative value of digital artefacts via sound digital forensics methods.

Further, the risks imposed by live examinations prior to forensic acquisition appeared not to be mitigated or addressed. While sometimes necessary, such examinations involve increased risk of data loss and modification. Key principles in digital forensics require that the 'acquisition process preserves a complete and accurate representation of the original data, [that] its authenticity and integrity can be validated' (Casey, 2007, p. 50)and to assure that 'when the method (or tool) used to gather and/or analyze digital evidence does change the original data set, the changes are identifiable' (Mocas, 2004, p. 67). The live examinations found in this study were insufficiently documented to satisfy these principles. A way to comply with the principles would be to justify the use of such methods and account for the possible and probable impact imposed on data integrity.

The documentation of acquisition tools was the best documented requirement (23%). Also, more than a half of the examination reports stated a mandate (60%). However, it was noticeable that in all processing stages there was a lack of data source, data set, and method description as well as examiner justification for the selected method and tool in relation to the forensic task. This suggests that "to some extend" technology-level documentation is preserved, while method and application level was missing. Moreover, with each subsequent processing step, the information become more scarce resulting in only three analysis reports found. These did not include any information about the analysis stage of the processing. The lack of analysis of potential digital evidence means that the relevance and probative strength of such evidence was not established according to accepted digital forensics methodology and tends towards hearsay.

Considering these findings, the following recommendations can improve the quality of digital forensics in Norwegian law enforcement:

- A formal process-level documentation based on reliability criteria and its integration and automation in the investigation procedure can increase accountability and trust in digital investigation and digital evidence. It falls outside the scope of this study to examine the concrete reasons for the identified lack of digital forensics processes and standards in investigations. However, the study developed and proposed a practical reliability validation method based on documenting concrete reliability criteria which can be used and extended as a template to create audit trails of digital forensics processing steps, and improve the reporting of digital forensics work performed by law enforcement.
- A digital forensics process and reporting system can help to create an audit trail, link together and document all the process steps and actions performed in relation to the relevant digital evidence found.
- Implementing and enforcing reliability validation procedures at technology, method, and application level in relation to the digital forensics employed can assist to improve the quality assurance and quality control in criminal investigations.
- The identified tool dependencies and reporting of tool results, rather than methodological application of digital forensics techniques points to a need for increased digital forensics competency, including better understanding of the process and its underlying principles for integrity preservation, digital chain of custody and chain of evidence, and process reliability assurance, especially with respect to examination and analysis of potential digital evidence.
- The dominant role of mobile phones in serious crimes investigations indicate that training and expertise in mobile forensics methods, tools and techniques should be a priority.
- The low number of analyses suggests a need to evaluate how and when analysis of potential digital evidence is requested and to assess if those responsible for making such requests possess the necessary level of competency.

Future research

This study shows a gap between the requirements of digital forensics as defined and described by the scientific community, and the way digital forensics are performed in practice by the Norwegian police. The developed reliability validation criteria can be used and improved in further research to examine if similar shortcomings are present in law enforcement agencies in other countries.

Further studies are necessary to understand the reasons behind this state of practice and the factors which impact the quality of digital forensics in law enforcement work. What constitutes a suitable reliability validation procedure in digital forensics investigations, necessary IT infrastructure, and approaches to integrate it in law enforcement daily work is also a subject of further research. In a broader interpretation of the results presented here, a study of how digital evidence is presented in court could further the understanding of the mutual impact between the digital forensics process, criminal investigation, and the overarching legal procedure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was made possible by financial support from the Norwegian Research Council (project 248094), Oslo Police District and Maria Sklodowska-Curie grand (Project ID: 722482).

Stig Andersen was employed at the Oslo Police District while conducing this study. The impartiality of his work is ensured by the Norwegian University of Science and Technology (NTNU), where he is enrolled as a PhD student.

Permission to access information about criminal investigation was granted by The Director of Public Prosecutions.[12] All personal information was removed prior to data processing according to procedures established in a Data Protection Impact Assessment (DPIA) conducted prior to data collection.

References

Alawadhi, I., Read, J., Marrington, A., Franqueira, V., 2015. Factors influencing digital forensic investigations: empirical evaluation of 12 Years of Dubai police cases. Journal of Digital Forensics, Security and LawVisited. https://doi.org/10.15394/jdfsl.2015.1207, 2021-02-25.

Andersen, S., Fossdal, J.P., 2021. Modelling criminal investigation: process, quality and requirements. In: Ijeh, A., Curran, K. (Eds.), Crime Science and Digital Forensics: A Holistic View. CRC Press/Taylor & Francis Group, pp. 41–66.

Ayers, R., Brothers, S., Jansen, W., 2014. Guidelines on Mobile Device Forensics. National Institute of Standards and Technology.

Brodeur, J.P., 2010. The Policing Web. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199740598.001.0001.

Casey, E., 2007. What does 'forensically sound' really mean? Digit. Invest. 4 (2), 49–50. https://doi.org/10.1016/j.diin.2007.05.001.

Casey, E., 2018. Clearly conveying digital forensic results. Digit. Invest. https://doi.org/10.1016/j.diin.2018.03.001.

Casey, E., 2019. The chequered past and risky future of digital forensics. Aust. J. Forensic Sci. 51 (6), 649–664. https://doi.org/10.1080/00450618.2018.1554090 publisher: Taylor & Francis _eprint:

Council of Europe, 2014. Electronic Evidence Guide: A Basic Guide for Police Officers. prosecutors and judges.

Council of Europe, 2021. Guide on article 6 of the european convention on human rights (criminal limb). updated on. https://www.echr.coe.int/documents/guide_art_6_criminal_eng.pdf. (Accessed 30 April 2021).

Custers, B., Vergouw, B., 2015. Promising policing technologies: experiences, obstacles and police needs regarding law enforcement technologies. Comput. Law Secur. Rep. 31 (4), 518–526. https://doi.org/10.1016/j.clsr.2015.05.005 visited 2021-06-21.

Erlandsen, T.E., 2019. Fallacies when Evaluating Digital Evidence Among Prosecutors in the Norwegian Police Service. Tech. rep., Norwegian University of Science and Technology.

European Network of Forensic Science Institutes (ENFSI), 2015. Best Practice Manual for forensic examination of digital technology. URL. https://enfsi.eu/wp-content/uploads/2016/09/1._forensic_examination_of_digital_technology_0.pdf.

INTERPOL, 2019. Global guidelines for digital forensics laboratories. https://www.interpol.int/en/content/download/13501/file/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory.pdf.

Gross, S., Mnookin, J., jan 2003. Expert Information and Expert Evidence: A Preliminary Taxonomy. Articles.

Heitmann, O., 2019. Digital Investigation: the Malnourished Child in the Norwegian Police Family? Tech. rep., Norwegian University of Science and Technology.

Horsman, G., 2018?show a. Framework for Reliable Experimental Design (FRED): a research framework to ensure the dependable interpretation of digital data for digital forensics. Comput. Secur. 73, 294–306. https://doi.org/10.1016/j.cose.2017.11.009.

Horsman, G., 2018?show b. 'I couldn't find it your honor, it mustn't be there!' - tool errors, tool limitations and user errors in digital forensics. Sci. Justice 58, 433–440. https://doi.org/10.1016/j.scijus.2018.04.001.

Hughes, N., Karabiyik, U., 2020. Towards Reliable Digital Forensics Investigations through Measurement Science. WIREs Forensic Science n/a (n/a). https://doi.org/10.1002/wfs2.1367 visited 2020-03-26.

Innes, M., 2003. Investigating Murder : Detective Work and the Police Response to Criminal Homicide. Oxford University Press, Oxford.

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2012. ISO/IEC 27037 eForensics Guidelines for identification, collection, acquisition and preservation of digital evidence. https://www.iso27001security.com/html/27037.html,visited,2020-09-03.

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2015. ISO/IEC 27042:2015 Information technology — Security techniques — guidelines for the analysis and interpretation of digital evidence. https://www.iso.org/obp/ui/#iso:std:iso-iec:27042:ed-1:v1:en,visited,2018-04-04.

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2021. ISO/IEC 27041:2015 guidance on assuring suitability and adequacy of incident investigative method. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/44/44405.html.

Jones, A., Vidalis, S., 2019. Rethinking digital forensics. Annal. Emerg. Technol. Comput. 3, 41–53. https://doi.org/10.33166/AETiC.2019.02.005.

Justis- og beredskapsdepartementet [Ministry of Justice and Public Security], 2012. Samfunnssikkerhet (Meld. St. 29 (2011-2012)) [White paper on social security in 2011.

Justis- og beredskapsdepartementet [Ministry of Justice and Public Security], 2020. Politimeldingen - et politi for fremtiden (Meld. St. 29 (2019 - 2020)). White paper on the Police - a police for the future in 2019]. https://www.regjeringen.no/no/dokumenter/meld.-st.-29-20192020/id2715224/.

Kohn, M., Eloff, M., Eloff, J., 2013. Integrated digital forensic process model. Comput. Secur. 38, 103–115 visited 2020-07-02.

Koper, C.S., Lum, C., Willis, J.J., 2014. Optimizing the use of technology in policing: results and implications from a multi-site study of the social, organizational, and behavioural aspects of implementing police technologies. Policing: J. Pol. Pract. 8 (2), 212–221. https://doi.org/10.1093/police/pau015 visited 2021-06-21.

Kumar, M., 2021. Solid State Drive Forensics Analysis—Challenges and Recommendations. Concurrency Computation. https://doi.org/10.1002/cpe.6442.

Marshall, A.M., Paige, R., 2018. Requirements in digital forensics method definition: observations from a UK study. Digit. Invest. 27, 23–29. https://doi.org/10.1016/j.diin.2018.09.004.

Mocas, S., 2004. Building theoretical underpinnings for digital forensics research. Digit. Invest. 1 (1), 61–68. https://doi.org/10.1016/j.diin.2003.12.004.

Montasari, R., Peltola, P., Evans, D., 2015. Integrated computer forensics investigation process model (ICFIPM) for computer crime investigations. In: Global Security, Safety and Sustainability: Tomorrow's Challenges of Cyber Security: 10th International Conference, ICGS3 2015, London, UK, September 15-17, 2015. Proceedings. Springer International Publishing, Cham, pp. 83–95. https://doi.org/10.1007/978-3-319-23276-8_8.

Nikkel, B., 2016. NVM express drives and digital forensics. Digit. Invest. 16, 38–45. https://doi.org/10.1016/j.diin.2016.01.001.

Pollitt, M., Casey, E., Jaquet-Chiffelle, D.-O., Gladyshev, P., jan 2018. A framework for harmonizing forensic science practices and digital/multimedia evidence, 2020-07-02. https://www.nist.gov/news-events/news/2018/01/framework-harmonizing-forensic-science-practices-and-digitalmultimedia,visited.

Riksadvokaten [The Director of Public Prosecutions], 2020. Mål Og Prioriteringer for Straffesaksbehandlingen I 2020 [Goals and Priorities for Criminal Proceedings in 2020]. Tech. rep., Riksadvokaten [The Director of Public Prosecutions.

Riksrevisjonen [Office of the Auditor General], 2021. Riksrevisjonens Undersøkelse Av Politiets Innsats Mot Kriminalitet Ved Bruk Av IKT [The Office of the Auditor General's Investigation of Police Efforts against Crime Using ICT]. Tech. Rep. Dokument 3:5, Riksrevisjonen [Office of the Auditor General.

Risinger, D., jun 2018. The five functions of forensic science and the validation issues they raise: a piece to incite discussion on validation. Seton Hall Law Rev. 48 (3).

Stelfox, P., 2009. Criminal Investigation : an Introduction to Principles and Practice,

[12] Ref. 2017/00668-010 IWI001639.2.

first ed. Willan Publishing, Devon, UK.
Stene, R.J., 2017. De nye kriminalstatistikkene [The new crime statistics]. https://www.ssb.no/sosiale-forhold-og-kriminalitet/artikler-og-publikasjoner/de-nye-kriminalstatistikkene.
Stoykova, R., 2021?show a. Digital evidence: unaddressed threats to fairness and the presumption of innocence. Comput. Law Secur. Rep. 42, 105575. https://doi.org/10.1016/j.clsr.2021.105575 visited 2021-09-03.
Stoykova, R., 2021?show b. The presumption of innocence as a source for universal rules on digital evidence — the guiding principle for digital forensics in producing digital evidence for criminal investigations. Comput. Law Rev. Int. 22 (3), 74–82. https://doi.org/10.9785/cri-2021-220303.
Stoykova, R., Franke, K., 2020. Standard representation for digital forensic processing. In: 2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering. SADFE, pp. 46–56. https://doi.org/10.1109/SADFE51007.2020.00014.
Sunde, N., Dror, I.E., 2021. A hierarchy of expert performance (HEP) applied to digital forensics: reliability and biasability in digital forensics decision making. Forensic Sci. Int.: Digit. Invest. 37, 301175. https://doi.org/10.1016/j.fsidi.2021.301175 visited 2021-06-22.
The UK National Police Chiefs Council, jul 2020. Digital Forensic Science Strategy. https://www.npcc.police.uk/Digital%20Forensic%20Science%20Strategy%202020.pdf.
Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., Watson, T., 2020. Quality standards for digital forensics: learning from experience in England & Wales. Forensic Sci. Int.: Digit. Invest. 32, 200905. https://doi.org/10.1016/j.fsidi.2020.200905.
Turnbull, B., Taylor, R., Blundell, B., 2009. The anatomy of electronic evidence. Quantitative analysis of police E-crime data. In: 2009 International Conference on Availability, Reliability and Security. IEEE, Fukuoka, Japan, pp. 143–149. https://doi.org/10.1109/ARES.2009.118 visited 2021-02-16.
United Nations Development Programme, 2020. (199 C.E.). Human development report : Norway. https://hdr.undp.org/sites/default/files/Country-Profiles/NOR.pdf.