NTNU
Norwegian University of
Science and Technology

Master's thesis

2019

Tor Stian Borhaug

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology and Electrical
Engineering
Department of Information Security and Communication
Technology

Master's thesis

Tor Stian Borhaug

The Paradox of Automation in Digital Forensics

June 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

The Paradox of Automation in Digital Forensics

Information Security: Digital Forensics and Cybercrime Investigation (MISEB)

Submission date:	June 2019
Supervisor:	Katrin Franke
Co-supervisor:	Rune Nordvik and Stewart Kowalski

Norwegian University of Science and Technology
Department of Information Security and Communication Technology

Abstract

Law enforcement agencies need to continually implement strategies and methods to meet modern technological demands. Increasing digitisation creates opportunities by enabling new methods, but also challenges that roots itself as an increasing backlog of criminal cases. The investigation process is ruled by fundamental principles that are absolute. Implementing methods that automates parts of the investigation process need to be balanced with these principles, due to the many pitfalls if not caution is applied when executed. Digital Forensics as a Service (DFaaS) is a system which uses modern technology to meet challenges created by the digitisation. This study looks at how law enforcement can use DFaaS in accordance with important investigation requirements and ultimately the rule of law.

A survey was developed and distributed to users of DFaaS in the Netherlands, which focused on the experiences with using such a system in digital investigations. It asked about the work processes in an investigation by applying a hybrid process model and several questions regarding the requirements of the investigation. The goal was then to use this data to do a socio-technical analysis, which is essentially assess the relationship between the social part of the system; its culture and structure, and the technical part; its methods and machines. The results indicate that several of the methods in the work process could be improved, and measures for this were proposed.

The importance of using a socio-technical system approach when implementing DFaaS was demonstrated. Without a stable structure with sufficient capacity, the efficiency of the system would suffer. Without the right competence and expertise in the system, the quality and the overarching principles are threatened. The Paradox of Automation becomes a reality if the efficiency is not balanced with the quality. Further research is required to develop a compatible process model, and verification procedures for DFaaS which seeks to balance these conflicting interests.

Sammendrag

Politimyndigheter må fortløpende implementere strategier og metoder for å møte moderne teknologiske utfordringer. Økende digitalisering skaper muligheter som muliggjør nye metoder, men også utfordringer som fører til økende straffesaksrestanser. Etterforskningsprosessen er styrt av overhengende prinsipper som er ufravikelige. Implementering av metoder som automatiserer deler av denne prosessen må balanseres med disse prinsippene, ellers risikerer man fallgruver ved at feil i etterforskingen kan oppstå. Digital Forensics as a Service (DFaaS) kan forklares på norsk som datatekniske undersøkelser som tilbys som en tjeneste. Dette er et system som bruker moderne teknologi for å få bukt med disse teknologiske utfordringene. Denne studien ser på hvordan politiet kan bruke dette systemet i harmoni med de overhengende kravene til etterforskingen, og til syvende og sist på en måte som ivaretar rettssikkerheten.

En spørreundersøkelse ble utviklet og distribuert til brukere av DFaaS i Nederland. Denne fokuserte på brukernes erfaring med bruken av et slikt system i den digitale etterforskningen. Det ble stilt spørsmål om arbeidsprosessene og kravene i etterforskingen ved hjelp av en hybrid prosess modell. Målet var å bruke responsene fra spørreundersøkelsen og foreta en sosio-teknisk analyse, som kort fortalt er en metode hvor man ser på forholdet mellom den sosiale delen av ett system; dette utgjør kulturen og strukturen, og den tekniske delen; dette utgjør metodene og maskinene. Resultatene fra dette tilsier at arbeidsprosessen i et slikt system kan forbedres, og tiltak for dette ble foreslått.

Det ble demonstrert viktigheten av at DFaaS blir implementert i et sosio-teknisk perspektiv. Effektiviteten i etterforskningen vil lide dersom man ikke har en stabil struktur med tilstrekkelig kapasitet til å serve brukerne. Videre vil de overhengende prinsippene og kvaliteten på arbeidet trues dersom den rette kompetansen og ekspertisen ikke integreres riktig. Automasjonsparadokset blir en realitet dersom effektiviteten ikke tilpasses riktig etter kvalitetskravene. Ytterligere forskning kreves for å utvikle en prosessmodell som er tilpasset DFaaS, samt metoder for å verifisere data og informasjon, slik at kravene for en effektiv og kvalitetssikret etterforsking innfris.

Preface

This dissertation marks the end of a three-year challenging, yet fulfilling and worthwhile endeavour. Having one foot in the practical field as a police officer, combined with the other foot in the research community have provided with both professional and personal insight. I think that by having an optimistic outlook while acknowledging the challenges law enforcement agencies face and experience, the potential pitfalls that could emerge while still having the utmost respect and sympathy for the men and women that try their best to meet multiple challenging demands, has motivated me on both a personal and professional level.

The fact that the Norwegian University of Science and Technology have created a joint Masters' thesis program together with the Norwegian Police University college is a step in the right direction. I am even more convinced now on the importance of merging science and theory with practise.

I want to thank my supervisor from NTNU, Professor Katrin Franke for her dedication, knowledge and support. My co-supervisors Stewart Kowalski and Rune Nordvik provided me with direction, ideas and inspiration which I appreciate.

I would like to thank my employer in the Norwegian law enforcement for the opportunity to combine studies with full-time work, which has been a struggle at times, but where my colleagues have demonstrated patience and support.

Together with myself on this ride, surviving the storm until the end was Odin Heitmann and Tom Erik Erlandsen. You provided fruitful discussions and made the effort much more fun. I also want to thank Stig Andersen and Grethe Østby for their input and experiences from their PhD studies. A big thanks also goes out to Aron for his positive humour, proofreading and attention to details.

Last, but not least, I show my deepest gratefulness to my wife which has shown patience and tolerance out of this world, supporting me and taking care of our little baby. When I am with you I feel I can conquer anything. And to my family that has not seen me much the last year; thank you for your patience and support.

Tor Stian Borhaug

Lørenskog, June 1st, 2019.

Glossary

Term	Definition
Anti-forensics	A general term for a set of techniques used as countermeasures to forensic analysis.
API[1]	A set of subroutine definitions, communication protocols, and tools for building software.
ASCII[1]	A character encoding standard for electronic communication.
Automation	Automation refers to a system or process that can operate without human intervention. Automation is best for repetitive, well-defined tasks. The less human intervention, the more efficient the Automation.
Binary	A numeral system that uses two symbols, typically "0" (zero) and "1" (one).
Bit	A binary digit often represented as multiple bits to measure units of information in computers.
Carving	A technique used in digital forensics to uncover data that has been previously deleted.
Encryption	Is a process of encoding a message or information in such a way that only authorized parties can access it.
Hexadecimal	A numeral system that uses 16 distinct symbols. Widely used as more human-friendly representations of binary-coded values.
I/O[1]	Is the communication between an information processing system, such as a computer, and the outside world, possibly a human or another information processing system.
IoT	Is the extension of Internet connectivity into physical devices and everyday objects. These are often called smart-devices.
IP[1]	The principal communications protocol in the Internet protocol suite for relaying datagrams across network boundaries. Often denominated as an IP address, which is a client or servers' address on a network.
LinkedIn	Is a business and employment-oriented service that operates via websites and mobile apps.
NAS[1]	A file-level computer data storage server connected to a computer network providing data access to a heterogeneous group of clients.
Operating System	System software that manages computer hardware and software resources and provides common services for computer programs. Examples are Windows, OSX or Linux.
Processing	A process is a set of activities that interact to produce a result. In this thesis it describes the act of going from raw data to more usable information by abstracting data to a higher more human-friendly level.
SQLite[1]	Is a widely used database management system.
Volatile data	A description of the degree to how long memory or storage units retain information.
Wearables	A form of IoT technology (see description over) that has a compact and practical size so that it could be wear on the human body. Smartwatches, heart sensors etc.
Write-blockers	A device that hinders write-operations on the source device, and only allows to read from it.

[1] See list of abbreviations

List of Abbreviations

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
DFaaS	Digital Forensics as a Service
GB	Giga Byte
I/O	Input/output
IaaS	Infrastructure as a Service
ICT	Information and Communication Technology
IDFPM	Integrated Digital Forensics Process Model
IoT	Internet of Things
IP	Internet Protocol
NAS	Network-attached Storage
NFI	Netherlands Forensic Institute
NIST	National Institute of Standards and Technology
NSRL	National Software Reference Library
OS	Operating System
PaaS	Platform as a Service
SaaS	Software as a Service
SQL	Structured Query Language
TB	Tera Byte
UTC	Coordinated Universal Time

List of figures

List of tables

1 Introduction

The first chapter provides an overview of the research questions and the motivation for choosing this topic. **Section 1.1** contains the backdrop to the research endeavour and its importance. **Section 1.2** contains the scope of the study and what it entails. Based on the scope, the research questions are formulated in **Section 1.3.** A summary of the methods used to answer these questions in given in **Section 1.4**. The target group for the study is depicted in **Section 1.5**. Clarification of the definitions used for the terms in this thesis is specified in **Section 1.6**, and last is the outline of the thesis given in **Section 1.7**.

1.1 Motivation and background

We are living in an increasing digitised era. Digitisation creates both opportunities and challenges for law enforcement agencies. Simson Garfinkel (1) wrote back in 2010 that the golden age of digital forensics was quickly coming to an end and that we were facing a digital forensics crisis. With the golden age he was referring to when the field of digital forensics started to grow in the early 2000's; when the suspect usually had just one computer, storage devices had standard interfaces, and recovering files was relatively easy. The coming crisis would be when storage devices are growing, when there are an increasing and diversified number of devices that often use flash storage and encryption, and there is an increasing use of cloud storage.

The characteristics of the digital forensics crisis the author would argue is today a fact, based on years of experience as a digital investigator myself.

Today the data is growing rapidly and the cost per byte is decreasing. By looking at the assortment at the biggest consumer webstore in Norway between 20 years back and today, for about 1000 Norwegian kroners (roughly 116 USD at April 2019) you could get a 4 terabyte (TB) single hard drive in 2019, a 1 TB hard drive in 2009 and a 4.3 gigabyte (GB) hard drive in 1999. The largest single drive that was sold had a capacity of 14 TB in 2019, 2 TB in 2009 and 37 GB in 2009 (2) (Figure 1).

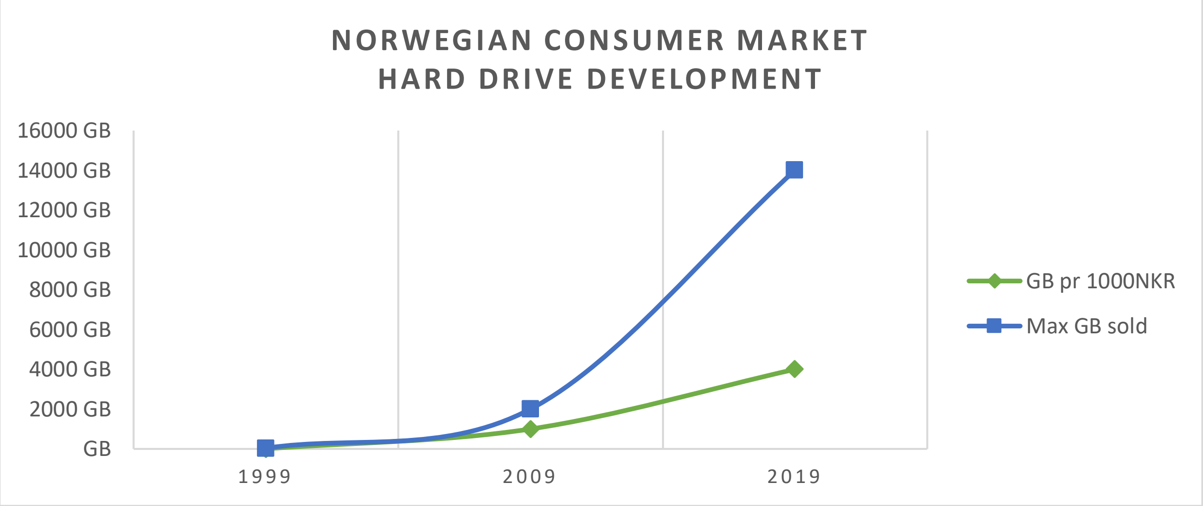


Figure 1: HDD capacity development

The development of storage has been given its own term called *Kryder's Law*. This states that the storage density of hard drives is on average doubling every 12 months. This is about twice the pace of *Moore's Law* which is the observation of an average doubling of the number of transistors on an integrated circuit every 18-24 months (3). This increase in storage density creates a major challenge for law enforcement agencies when investigating crimes where electronic devices and other digital traces are used for evidence, especially since today almost every criminal case has some form of digital data associated with the crime. Considering that the time it takes to process the data is dependent on the processing power, which development is currently doubling every 18-24 months and the storage density is doubling every 12 months there is a systemic gap that will continue to grow.

This growing gap and the massive amounts of data that are produced through our day to day activities has been given its own term called *big data*. Big data has been listed as the top issue facing forensic professionals towards 2019 (4).

The increase in the number of devices that are seized for analysis, the number of cases where digital evidence is crucial, and the volume of potentially evidence-rich data stored on each item seized, causes increasing case backlogs (3). This leads to investigation delays, case closures and possible crucial evidence getting lost or missed which cause a bad quality overall on the investigation. The existing forensic software solutions are beginning to address scalability issues (3) and to keep up with this development there is a call for new methods, one being more automation of tedious tasks, but also more effective ways to carry out the digital investigation.

Scenario, part 1

Harry, a police detective working in Torskevik[2] Norway was contacted by the national Cyber Task Unit[3] (CTU) informing him that a person (hereby referred to as the subject) with residence in Torskevik had been downloading images and videos that was depicting sexual abuse of children. The local Digital Forensics Unit had assisted with the seizure of four computers, a network-attached storage (NAS) unit with multiple hard drives attached and two mobile phones. The devices had been acquired as forensic disk images and stored at the police server in Torskevik. The subject also had an online cloud storage that was acquired.

Unfortunately, the local Digital Forensics Unit did not have the capacity to process the images, because of a large backlog of cases and another high-profile case that came right after this case. It had crippled the processing capacity and made the digital forensics experts occupied for several months ahead.

The subject had a high position in the local municipality and rumours started to spread quickly in the local society, with a lot of exaggeration. The subject claimed he did not know why his Internet Protocol (IP) address was recorded downloading illegal data material, and the accusation was totally incomprehensible to him. The case was only based on a hypothesis, and so far, nothing had been confirmed. Harry and his colleagues started to feel the pressure to give some answers.

[2] Not a real place, made up for the story

[3] Not a real unit, made up for the story

Harry had learned at the police academy where he graduated two years prior, that one should avoid going through devices manually because this could, at worst, change or delete data and render it useless in court. But there were not many options right now.

The need for technical solutions that supports the human examiner to keep up was identified by Franke and Srihari (5). The use of what is referred to as *computational forensics,* is a way to associate human expertise with modern technology. This provides methods to better analyse evidence by overcoming limitations of the human cognitive abilities, and it has a scientific basis for methods and procedures to analyse large volumes of data which are not humanly possible. These methods represent human expert knowledge by implementing recognition and reasoning abilities in machines. One approach to be used that have the potential to implements such abilities is the Digital Forensics as a Service model.

Digital Forensics as a Service (DFaaS) is a service-based approach for processing and investigating high volumes of seized digital material where the data is sent to a centralized system that automatically extracts traces from the data and gives digital investigators, detectives and analysts access to the traces (6). The law enforcement in the Netherlands have used this approach since 2010 and they have seen that the backlogs have reduced after implementing this (6).

A highly automated system like DFaaS has a lot of benefits in that it will make the law enforcement more effective in digital investigations. But with more automation there are often new challenges that are introduced. One way to look at this, is through what is called the *paradox of automation* where the more efficient the automated system, the more crucial the human contribution of the operators. Humans are less involved, but their involvement becomes more critical (7). This creates a dilemma in that law enforcement agencies need to automate more (to reduce backlogs etc.), but with more automation the risk for miscarriage of justice is increased if the human control is not equally strengthened. If we do not automate more, we also risk miscarriage of justice, with increasing backlogs, case delays and potential evidence getting deleted.

Traditionally, the digital investigator has been the expert resource in the digital investigation. Further, they have had the required knowledge to control the quality of the data being investigated and maintaining supervision of the digital forensics process. This thesis looks at the relationship between the human role and modern technology, in this case DFaaS, as a potential aid for the coming challenges the law enforcement agencies face in their investigations. For the investigation process to be in line with the rule of law there are rules or principles that must be followed. Will the automation that this model provides support these rules and principles so that it is in line with providing for efficiency in the investigation, with the quality in focus?

Scenario, part 2

Harry called the CTU and explained the situation. He got some exiting news that CTU was doing a pilot and testing out a new system that would make it possible to process large amounts of data in a short time and make the content available for the detectives at remote locations in Norway. CTU thought Torskevik would be a great test for this system since the location of the police unit was the most distant to CTU, far up north in Norway. Harry got the offer and gladly accepted it.

Harry asked the local Digital Forensics Unit to send the disk images to CTU. Because of the amount of data, this had to be physically shipped. A couple of days went by to copy out the data to physical hard drives, and it was then shipped to CTU. Harry got feedback from the CTU the next day that the data had triggered the hash database of known child sexual abuse images, as well as the pattern recognition had categorised multiple images to be suspicious and possible sexual abuse images as well. The data would take some time to process because of the large amount, but he got instructions for logging into the DFaaS system with direct access to the data to do a preliminary review.

(Continues on page 51)

1.2 Scope of the study

The focus of the thesis is on the people involved in the investigation (digital investigators and detectives) using DFaaS. How is the compatibility between users and methods in the digital investigation process when using DFaaS, and how does this relate to the principles of the digital investigation?

The main goal is to see how the law enforcement can use DFaaS to meet the modern technological demands of increasing digitisation and more data everywhere with focus on what is required in an investigation, which is the quality and efficiency requirement.

The data that will be collected is based on the perception of the digital investigators and detectives themselves. The focus is not on privacy and security issues, but on work processes in the investigation matched against the investigation requirements.

Second, the study is focusing on the digital investigation process, not the preparation/planning-, incident response- or the presentation phase. This is congruent with systems thinking where the system in focus is defined.

The study is not meant to be a review of a concrete DFaaS tool, but more an exploration of DFaaS in general. It will not be provided for a technical review of DFaaS. How a system like this should be built in terms of hardware and software design is not in the scope of the study. The study is largely sociological with the focus on people (users of DFaaS) and methods (processes of the DFaaS model).

The focus-subjects of the thesis is shown in the circle in Figure 2. The subjects outside of the scope (delimitations) are placed outside the circle to demonstrate their peripheral roles, which means they are still relevant, but not part of the scope for this thesis.

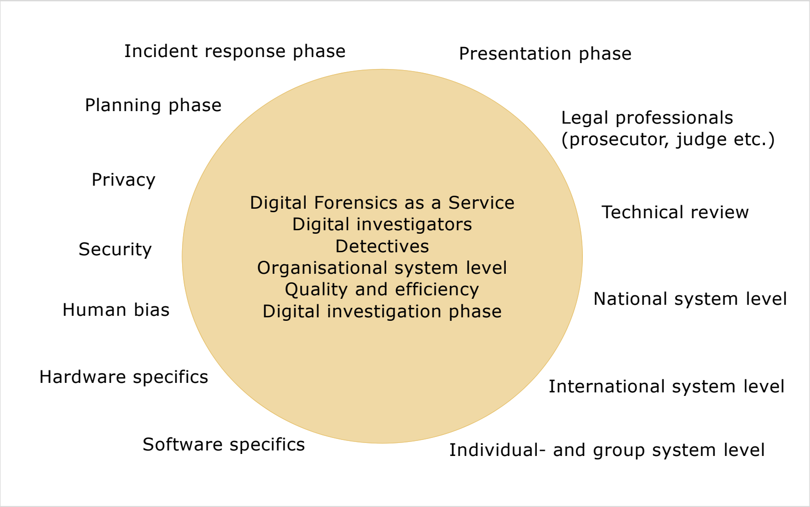


Figure 2: Scope of the study

1.3 Research questions

Based on the background, motivation, and scope of the study; the research questions are introduced.

Main problem

1. How can law enforcement agencies use Digital Forensics as a Service to meet the modern digitisation challenges with focus on fundamental investigation requirements?

Since the research question is very general, not affecting its importance, some sub questions are formulated. The goal of asking these questions is to deconstruct the main problem into smaller more manageable problems, which together makes the foundation to answer the main problem.

Sub questions

1.1 Are the digital forensics process models valid work models for DFaaS?
1.2 What socio-technical measures are suitable to attain efficiency in DFaaS?
1.3 What socio-technical measures are suitable to attain quality in DFaaS?
1.4 How can the socio-technical system in DFaaS be balanced and what constitutes socio-technical balance?

The objective of the first question is to look at the process, which serves as a guideline for the investigation. Then the two fundamental investigation requirements *efficiency* and *quality* are investigated independently. The last question seeks to tie these requirements together.

In the figure below, the research question is illustrated. The flow diagram starts with the underlying problem caused by the *increasing digitisation*. This generates both *opportunities* and *challenges*. *Law enforcement agencies* need to take advantage of the opportunities and mitigate the challenges to meet the *fundamental investigation requirements*. *Digital Forensics as a Service* can be described as a "toolbox" that law enforcement agencies can use, but with new tools, new *opportunities* and *challenges* are created. How can law enforcement balance these opportunities and challenges?

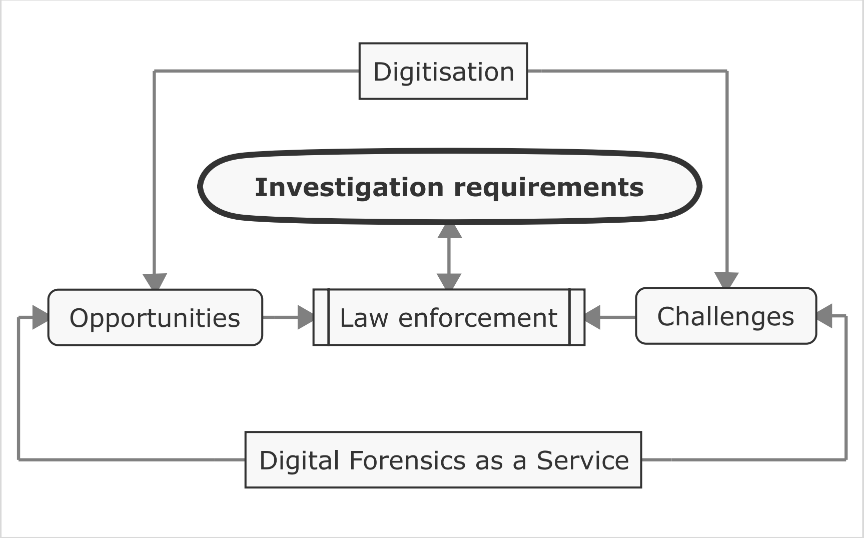


Figure 3: Research topics and problem

1.4 Research method in brief

To answer the research questions, it would be helpful to have some ways of measuring the different parts of the DFaaS model and its interrelationships. For this task, the socio-technical systems model will be used. This is often used to analyse aspects of implementing for example; new technology in relation to the compatibility with human usage (8). The model is split into a *technical* side and a *social* side. The technical aspects of a system are the *machines* and *methods*. The social side of the system includes the *culture* and the *structure*.

Digital Forensics as a Service is in this thesis split into a technical side and a social side. There is probably an application that is used, this constitutes the *machines*. The *methods* used for the investigation is described in the digital investigation process model. The people involved in the investigation embodies the *culture* and the law enforcement organization and infrastructure is the *structure* of the socio-technical system.

The point is to see the relation and the compatibility between these, and if this is congruent with a socio-technical system in balance. Figure 4 illustrates the interrelationship in the system, and the arrows shows that all the elements of the model (methods, machines, culture and structure) are in relationship and affects one another. One example is if we introduce a new machine that the people in the system have no competence of handling. This makes the system imbalanced by having either a weak link between the technical and social side, or too much weight on one of the sides. This could tip the scale and have fatal consequences.

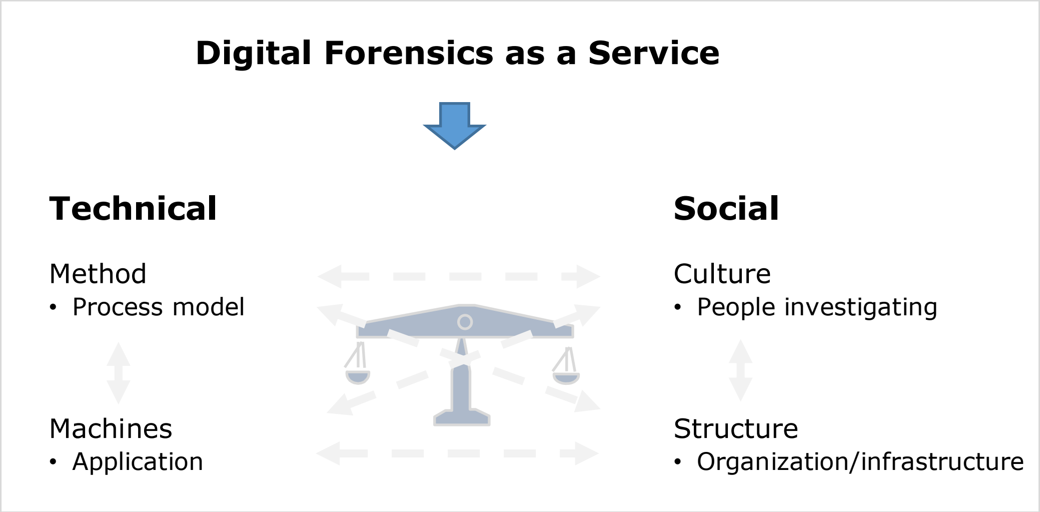


Figure 4: Socio-technical systems model, based on Kowalski (8)

The goal of the research endeavour is to look at the interrelationships between the different parts of the system to see how the investigation quality can be uphold to a high standard. The exact data sources that represents the processes, application, people or organization, will be identified in Chapter 3.

1.5 Target group

The target group is the research community doing similar research and law enforcement agencies in general, especially the decision makers and those involved in digital investigations. Other decision makers on other system levels (national etc.) will most likely be in this group as well.

1.6 Clarifying terms

Some of the terms that are used throughout the thesis, should be clearly defined to avoid misconception or confusion.

Detective[4]: Will be used to describe the role of the person in the digital investigation with more tactical case experience, and not necessarily with any technical competence. Other terms with similar meaning would be case investigator, police sergeant etc.

Digital forensics[5]: The term *digital forensics* will be referred to as a general description of the method used to investigate, which is the application of forensic science to digital information.

Digital investigation[5]: The term *digital investigation* will be used as a general description of performing investigations in the digital domain. Other terms with similar meaning are for instance; digital forensics investigation and cybercrime investigation. This term, together with digital forensics, is directly related to the process, which is reviewed in Section 2.2.1. For clarification, a concrete process model is used in the experiment and discussion of the study.

Digital investigator[4]: Will be used to describe the role of the person with more technical knowledge and competence within the field of digital investigation and digital forensics. This role could also be called for example; digital forensics investigator, digital forensics examiner or subject expert (10).

Fundamental investigation requirements: These refer to the *efficiency* and *quality* in an investigation. For law enforcement agencies to handle the increasing digitisation and resulting backlogs there will need to be efficiency. To do this in accordance with the investigation principles such as evidence integrity and chain of custody there will need to be quality (refer to Section 2.2.2 for a detailed description of these principles).

- **Efficiency:** The term *efficiency* is used as a measure to describe how the investigation reveals relevant traces in an efficacious and expedient way, in general to speed up the investigation not affected by the amount of seized data in the case.
- **Quality:** The *quality* term in this regard specifies an investigation that is congruent with its principles and ultimately with the rule of law.

Hansken/Xiraf: Name of the DFaaS system developed by the Netherlands Forensic Institute. Hansken and Xiraf may be used interchangeably, but they are basically the same, Xiraf is just the name of the previous version of Hansken.

[4] This is the same definitions used by van Baar et al. (6)

[5] This is the same definitions used by Årnes (9)

1.7 Thesis outline

Chapter 1: The first chapter provides an overview of the research questions and the motivation for selecting the topic.

Chapter 2: Provides the theoretical foundations of the thesis. What creates backlogs and what are ways to mitigate this? Some of the methods to meet these challenges are reviewed, and the core of the thesis, namely the digital investigation process is being defined. What principles governs the field, and how can errors occur.

Chapter 3: Gives the reader an insight into the methods that were used to gather data and examine the results of this.

Chapter 4: Focuses on the output from the data gathered (results) and does an analysis of the results.

Chapter 5: Addresses the depths of the results presented in Chapter 4. What does the data tell us? The sub questions are restated and discussed further.

Chapter 6: Last chapter brings a conclusion and what this implies for the field of study and practise. Then suggestions for further work are proposed.

Scenario
A four-part scenario is used throughout the thesis (Chapter 1, 5 and 6) to exemplify some of the topics of the thesis:

- Part 1 (p.2): Challenges with the modern digitisation.
- Part 2 (p.3): Opportunities this provides in form of new technologies/methods.
- Part 3, (p.51): Errors due to insufficient competence and bad implementation.
- Part 4, (p.71): Solution with the right competence and cooperation.

2 Theoretical foundations

In this part of the thesis, a recursive review of the relevant topics that was introduced in the first part in the study will be provided. The inverted triangle approach (11) will be used, starting with the general content associated with the topic moving towards the specific information directly related to the project.

Section 2.1 investigates the characteristics of the technology evolution and how is this a challenge for the law enforcement. **Section 2.2** provides definitions of the digital investigation and starts off by looking at the criminal case from a high level, then dives into the field of digital forensics by going into depth on the processes and looking at related work and the development of the process models. What principles governs the field that is used to mitigate for errors. **Section 2.3** looks into concepts that can explain how errors can occur, and the paradoxical situation that may arise when implementing new technology such as automation. **Section 2.4** reviews the current state of the art along with a summary of the chapter.

2.1 Underlying challenges

2.1.1 Big data and increasing digitisation

The term big data is often used as an umbrella concept that describes the development of an increase in the amount and variety of data sources, and the increasing volume of data overall. The following description is taken from Gartner and complies with this definition:

> Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (12).

The development in technology has led to a dramatic drop in storage device prices, an increase in magnetic storage density and diffusion of solid-state media, increasing amount of personal mobile devices, increase in the adoption of cloud services, network traffic growing in size and speed and "everything" getting connected to the internet (13).

These facts and their implications for the law enforcement, are that almost every criminal case involves some form of digital investigation. The number of digital devices in each case is increasing and the storage volume of each device is growing. The diversity and complexity are increasing with different storage mediums, file formats, file systems, cloud storage, and IoT devices or smart devices such as wearables, cars and even components of people's homes like refrigerators and alarm systems. On top of this there are the challenge of anti-forensics (e.g., volatile-, obfuscated-, hidden- or encrypted data) which could be intentional or not. One of the consequences of this is that many agencies are confronted with increasingly larger backlog of criminal cases (3, 14, 15).

The ability to "keep up" with the developments for the law enforcement is nothing new, it was pinpointed as a major dilemma back in 2001 with the rapid changes in digital technology (16).

Today's tools that are utilized in a digital investigation like Encase (17), Axiom (18), UFED (19) and FTK (20) are convenient and user-friendly, and they provide investigators the ability to browse file systems, conduct keyword searches and employ a range of other analysis techniques. But these tools, that Ayers refer to as "first generation forensic tools" are struggling to keep pace with the modern analysis workloads (21).

There is also important to mention that the increase in backlogs are not just purely a technical issue. According to Casey et al. (22) one issue is the expectation of other investigators, attorneys and judges in that it is expected an in-depth investigation and analysis of every item. It is also important to scale the examination to the investigation of the case, by having a better collaboration where every party understand the full life cycle of a case involving digital evidence.

Analysis of items in a digital investigation is often a very time-consuming matter. First, the data must be processed with the never-ending battle regarding storage and performance bottlenecks. Investigators need to analyse computer systems that could recently just have been purchased by the subject. The law enforcement is required to have updated components in their workstations, and thus using top-of-the-line systems to analyse other top-of-the-line systems. Data from these systems have to be analysed in a matter of hours or days, what a subject spent weeks, months, or even years assembling (23). In addition to being time-consuming, it requires attention to details and an in depth knowledge of how computers works, and the level of expertise needed to understand such as compiled object code in reverse engineering is quite high (24).

With an increasing backlog, investigators are pressured to work faster. This potentially leads to less time to study underlying concepts and getting up to date on topics that could improve their processes, both in terms of effectivity and quality (25). The legal implications of growing backlogs are suppression of evidence due to the delay of examination of digital evidence, delay in prosecution which has the potential to provide for more time and opportunity to commit additional offenses, and likelihood that evidence in less serious cases will be skipped over for more high profile cases and driving up the bar that must be reached to consider a case worthy of prosecution (26).

In Norway, there are specific demands for the criminal proceedings in terms of minimizing the time period from the accuse to the final verdict. This is specified in the European Convention on Human Rights. Further, the Criminal Procedure Act § 226 points out that the investigation should be carried out as soon as possible, so that no one are being falsely accused or subject to disadvantage. The question of an indictment should be carried out as soon as the case is adequately prepared according to the Criminal Procedure Act § 249. The respect to those involved (accused, victims etc.) also requires an effective investigation (27).

2.1.2 The backlog problem

Many digital forensics laboratories have large backlogs. It is not unusual that these range from six months to one year (22). According to a report from the Irish national police in 2015 the delays in the digital forensic investigations has been up to four years, which is mainly due to the backlogs of cases (15). Gogolin surveyed the Michigan Sheriff Departments and found that many law enforcement agencies had a digital component in 50% or more of their cases and the backlogs was exceeding 2 years (28).

Audits of the FBI have also shown large amounts of backlogs: "An audit report of the Office of the Inspector General U.S. Department of Justice highlights that a backlog of 1566 outstanding cases existed, 57% of which had waited between 91 days to over 2 years" *(29, p. 2, reffering to and audit of FBI's Philadelphia regional computer forensic laboratory, Radnor, Pennsylvania in 2015)*.

The author contacted the National Police Directorate in Norway to request for similar investigations on the amount of backlogs with regards to digital evidence in Norway, but there was no statistics regarding this phenomenon in Norway. The same was done towards the Netherland's authorities on statistics; Centraal Bureau voor de Statistiek and the Research and Documentation Centre at the Ministry of Justice and Security, but there was no such statistics there as well. The intention was to compare and see if the introduction of DFaaS (in the Netherlands) have had any impact on the backlogs in terms of concrete numbers, but unfortunately this was not possible due to their response. However, according to some articles from the Netherlands Forensic Institute the backlogs have indeed decreased and the agencies have come back in line with the investigations after the implementation of DFaaS (6, 30, 31).

The most compelling evidence that was found regarding the backlog problem in Norway was in the STRASAK (criminal case statistics) reports for the years 2015, 2017 and 2018 (32-34).

> The reason for the increase is probable due to that several police districts have investigated serious internet-related sexual abuse criminal cases where the amount of seized digital evidence have been large. The increase in the internet connectivity speed and greater capacity of each storage device allows police seizure of abusive material (images and videos containing sexual abuse against children) which has increased significantly in recent years (34, p. 59).

The Norway's public inquiry to the Government in 2015 also states that there can be large backlogs related to electronic traces within all forms of crime (35, p. 265).

2.2 Digital investigation fundamentals

"The goal of any investigation is to uncover and present the truth… This goal is the same for all forms of investigation whether it be in pursuit of a murderer in the physical world or trying to track a computer intruder online." (36, p. 187).

Figure 5 illustrates how a criminal case evolves, starting with an incident that is detected and identified. The investigation, digital or non-digital, lays the foundation for a possible indictment or dismissal. If indictment, the case is prosecuted in court that leads to some decision or appeal (10), but these are topics outside the scope of the thesis.

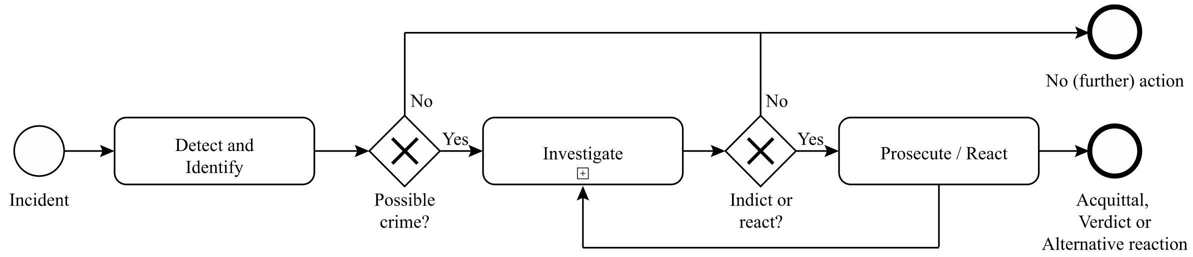


Figure 5: Criminal case process model (10)

The investigation is performed to determine if a crime has occurred, and to gather data and information to shed light on the incident. The goal is then to find out what happened, where, when, who, why, to whom, and how it happen (10).

Figure 6 is a more detailed look at the investigate phase shown in Figure 5. The investigation process involves; formulate possible hypotheses, identify required information to evaluate hypotheses, collect and process data and evaluate the hypotheses in light of the collected information (10).

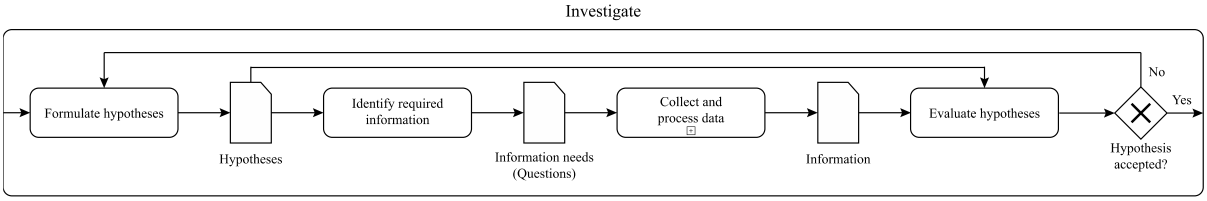


Figure 6: Criminal investigation process model (10)

There are two main types of digital-/cyber-crimes according to McGuire and Dowling (Cited in: 37): Cyber-enabled and cyber-dependent crimes. The first is traditional crimes which use Information and Communication Technology (ICT) to increase their scale or reach, while the second is crimes committed only using ICT.

Reyes et al. (Cited in: 37) defines it further into four different categories:

- ICT as a target
- ICT as a tool
- ICT affiliated with a crime
- Crimes against the ICT industry

A central part of the digital investigation is the application of digital forensics. It could be compared to the crime scene forensics in the physical world. Digital forensics is based on forensic science which is the application of scientific techniques and theories to law and is ultimately tested in court. Forensic science can help reconstruct crimes and generate leads. In digital forensics the scientific method is applied to analyse evidence, reconstructing the crime and testing hypotheses (36). One of the earliest definitions of the digital forensics science was made at the first digital forensic research workshop (DFRWS):

> The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations (16, p. 16).

The digital investigation should be an integrated part of the whole investigation, which also includes the tactical investigation (interviews, seizures of physical objects etc.). The actors in the digital investigation is normally not just the digital forensics experts, but also detectives with no expert knowledge at digital forensics. The end goal of most digital investigation is to identify a person who is responsible and therefore the digital investigation needs to be tied to a physical investigation (38).

If digital forensics is the application of scientifically derived and proven methods to digital information, then this needs to be forensically sound to trust the process. An investigation is *forensically sound* if it adheres to the digital forensics *principles* and *process* (9). The process is built on the theoretical framework of process models, and the principles are based on the *integrity* of the evidence and the *chain of custody* of the whole process from a to z.

We want to alter the original evidence as little as possible and any changes should be documented and assessed throughout the process to reach the required integrity requirements. There is a potential for changes to occur on the investigated device even if we are for example applying measures such as write-blockers, and most certainly when we are acquiring live systems, which is often the only option when dealing with encryption and volatile data. If the acquisition process preserves a complete and accurate representation of the original data, and its authenticity and integrity can be validated, it is generally considered forensically sound (36).

The next sections will focus on what makes the investigation of a high-quality standard.

2.2.1 Overview and development of the process models

Many process models have been developed through the years. These have sought to systematize and formalize the digital investigation process. Casey writes that the motivations for developing process models are numerous:

> Such process models serve as useful points of reference for reflecting on the state and nature of the field, as a framework for training and directing research, and for benchmarking performance against generally accepted practice. Using a formalized methodology encourages a complete, rigorous investigation, ensures proper evidence handling, and reduces the chance of mistakes created by preconceived theories, time pressures, and other potential pitfalls. Another purpose of these models is to refine our understanding of what is required to complete a comprehensive and successful investigation in a way that is independent of a particular technology in corporate, military, and law enforcement environments. An effective process model identifies the necessary steps to achieve goals, and can be applied to new technologies that become a source of digital evidence (36, p. 188).

The process models should enhance quality of the digital investigation and provide for some standardisation of the process. Numerous process models have been developed through the years, and many of them have similar steps that seeks to guide the investigation. Selamat et al. looked at 13 of the process models that was developed up until 2008 and they constructed five phase terms that summarized the models and to which they used to map the processes to. These phases were:

1. Preparation (plan, strategy, legal support).
2. Collection and Preservation (locate, collect and preserve evidence).
3. Examination and Analysis (investigate, validate, interpret and discover data/test hypotheses and draw conclusions).
4. Presentation and Reporting (clarifying the evidence and document findings).
5. Disseminating the case (return evidence, review process, preserve knowledge).

They discovered that the phases that most of the frameworks consisted of was phase 2 – collection and preservation, phase 3 – examination and analysis, and phase 4 – presentation and reporting. But they also argued that phase 1 – preparation, and phase 5 – disseminating the case was important to ensure completeness of the investigation (39).

Abulaish et al. did a similar comparison in 2018 and mapped all the sub-processes of the different process models into seven phases or stages. The most occurring sub-process for each phase was: Identification, preparation, preservation, collection, analysis, presentation and review (40).

Casey also compared different models (36) and did a similar mapping, but with a couple of important differences. First, preparation (getting a plan of action to conduct an effective digital investigation and obtaining supporting resources and materials) came before identification (finding potential sources of evidence). Then came preservation which also included collection as a sub-step (preventing changes of digital evidence and collecting data). The phases examination (the process of extracting and viewing information from the evidence and making it available for analysis) and analysis (the application of the scientific method and critical thinking to address the fundamental questions in an investigation) was closely related and placed at the same stage. And last was the presentation step (reporting/presenting of findings).

The development of the process models can be categorised into three phases: The earliest process models which tried to define the entire digital forensics investigation, while the next ones centred around specific use cases (such as cybercrime cases or triage) or particular steps in the investigation process (such as collection, examination or analysis). The more recent ones have focused on emerging trends, problems or methods (such as cloud, data reduction/mining, IoT or field processing) (14).

The technological development in terms of devices being investigated and tools being available to investigators have called for a dynamic development in the methods used. There is not one model that fits all digital investigations. Some a very specific scenario, while others can be applied to a vast scope of case types. Some are very general and others more detailed. The process models should serve digital investigations, not dictate them, and they can be useful under certain circumstances, but have limitations under others (36, 39-41).

One of the challenges with using the frameworks have been that the terminologies used for the processes have not been accordingly standardized and differing terms that refers to the same process or steps have been used (42).

In the integrated digital forensic process model (42) the authors derived a proposed solution that integrated the different phases of six previous models and purified the terminology used. They searched for similar meaning in the terminology to reduce the number of required processes and established the following processes; preparation, incident, incident response, physical investigation, digital forensic investigation and presentation.

The digital forensic investigation process was split into 17 very concrete tasks, and the physical investigation process occurs, according to the authors, in parallel with the digital investigation if the crime is not isolated to the digital space (see Figure 7).

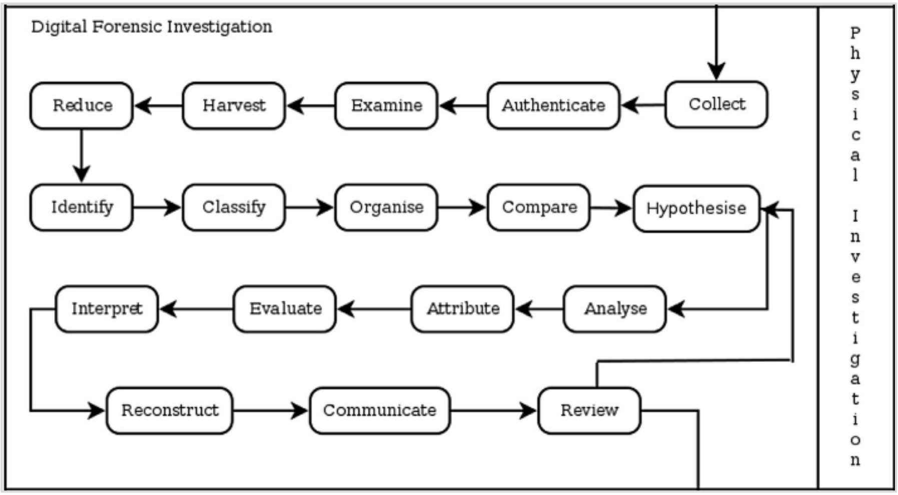


Figure 7: The Integrated digital forensics process model (42)

This phase was depicted as the data collection and processing phase seen in the in the Data Collection and Processing process model (10) (see Figure 8). This process model has the following steps; identify, locate and acquire data sources, acquire data and traces, explore and examine data and traces, and analyse data and traces. The model also shows the abstraction level of data from source/item, to raw data/trace, to significant data. The author points out that this phase is performed by a subject expert, i.e., a person with the necessary knowledge and skills.

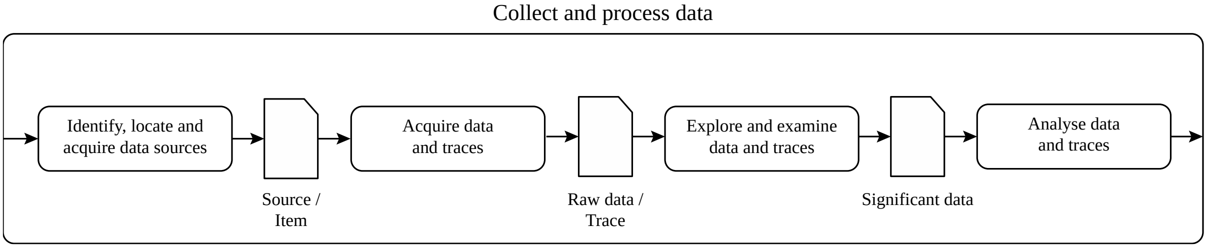


Figure 8: Data collection and processing process model (10)

Several of the process models points out that the different stages or steps should not be linear, but rather an iterative process (36, 43-45). This could be the case if new evidence is identified in the analysis process, then one would return to a process of preparation, evaluation, identification, collection, preservation, and organization of the new device and data (45). These are referred to as first-tier phases by Beebe and Clark (44), which when started should be sequential and non-iterative to the maximum extent possible.

The sub-phases are more iterative in nature, which the authors in (44) define for the analysis phase as; survey, extract and examine (SEE). Survey is when one gets an overview of the device, like mapping of file systems, disk partitioning etc. In the extraction phase techniques such as keyword searches, filtering and pattern matching are used to prepare data for examination to achieve confirmatory and/or event reconstruction goals.

Regarding the analysis phase it is important to make a distinction regarding what kind of competence is required. Casey (46) used the terms *investigative activities*, *technical processes* and *evidence evaluation* to differentiate the phases of digital investigation. In investigative activities technical knowledge is not required for developing leads in an investigation and to find evidence sources, suspects and victims. Technical processes involve extracting and observing data from digital devices, which is a technical task that does not require interpretation or evaluation. Evidence evaluation on the other hand involves interpretation and evaluation of digital evidence and requires more specialized knowledge. Making this extinction helps avoid problems associated with unqualified individuals attempting to evaluate digital evidence without the necessary expertise according to Casey.

Sunde (47) extended this research and refined the definitions on what she called the analysis sub-phases with a clear distinction on what was required in terms of technical competence and investigative competence. Three sub-phases were described. The first, *technical analysis* means to examine, verify and evaluate the quality of technical data that contains relevant information to the case. This requires pure technological competence. The second, *content analysis* is done by identifying and documenting relevant content information from electronically stored data, which requires purely investigative competence. The third, *digital evidence evaluation* is done when trying to determine accuracy, causation, linkages, spoliation and meaning, and requires both technical and investigative capabilities (47).

2.2.2 Principles

Årnes listed two fundamental principles of the digital forensics investigation. This was *evidence integrity* and *chain of custody*:

> Evidence integrity refers to the preservation of evidence in its original form.
>
> Chain of custody refers to the documentation of evidence acquisition, control, analysis and disposition of physical and electronic evidence.
>
> (9, p. 22)

Digital evidence is often highly volatile, and it can easily be altered or deleted without a trace throughout the investigation process. It can change both while within a computer and while being transmitted (e.g., during the process of evidence collection). But it can also be duplicated exactly, so it is possible to examine a copy without the risk of damaging or altering the original (48).

The method of upholding evidence integrity in the data is mostly done by using write-blockers and applying cryptographic hash calculations to detect possible alterations in the evidence through comparison. And since most of the digital forensics process is tool based, there must be some legal standard that these tools must pass to be forensically sound. The Best Evidence Rule from the US Federal Rules of Evidence points out that the information shown to the court must reflect data accurately and for most purposes it is sufficient to show that a tool does not alter the evidence, and that the results could be replicated (49).

Replicability can further be split into either repeatability or reproducibility, which are requirements set by the National Institute of Standards and Technology (NIST) for the test results to be considered as admissible evidence. Repeatability refers to "obtaining the same results when using 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 refers to "obtaining the same results being obtained when using the same method on identical test items in different laboratories with different operators utilizing different equipment" (50).

To fulfil the requirements of replicability and consistency for the scientific process, the industry has adopted multiple good practices, processes and procedures. There are also numerous standards that have been introduced for the forensic laboratories, such as ISO 17025 (general requirements for the forensic laboratories), ISO 31000 (risk management) and ISO 9000 (quality management) (48).

The Daubert standard (Daubert v. Merrell Dow Pharmaceuticals Inc, 1993) is used in most U.S. states and provides judges with an objective set of guidelines for accepting scientific evidence in court and consists of the following criteria's:

The method or procedure;

- must have been independently tested,
- must have been published and subjected to peer-review,
- must have its known or potential error rates available,
- must have standards and controls concerning its operation in existence,
- must have a general acceptance in the community that uses it. (36, 48, 50-52).

Since the tools used can have bugs that could lead to incorrect or incomplete findings, it is important to test the reliability of the tools. This will reduce the risk of mistakes, misinterpretations, missed evidence, and potential miscarriage of justice (36).

The ISO 27037 standard (*"Guidelines for identification, collection, acquisition and preservation of digital evidence"*) describes three main governance principles for digital evidence: Relevance, reliability and sufficiency. Some requirements for the handling of digital evidence are also given: Auditability, justifiability and either repeatability or reproducibility (9).

Examples of more practical guidelines or best-practices are the U.S. Department of Justice, National Institute of Justice's (NIJ) "Forensic Examination of Digital Evidence: A Guide for Law Enforcement" (53) and the "Good Practice Guide for Digital Evidence" from the Association of Chief Police Officers in the U.K. (ACPO) (54).

NIJ's guide are intended for members of the law enforcement who are responsible for the examination of digital evidence and because of the complex issues associated with digital evidence examination they recognized that "it (the guide) may not be feasible in all circumstances", and that it should act as a suggestion for the steps taken during the digital investigation process, rather than a mandate. It is split into five chapters with suggested procedures and a defined principle for each chapter. The summary of these principles are:

- The field demands specially trained personnel, support from management and sufficient funding.
- The digital evidence should be thoroughly assessed with respect to the scope of the case to determine the course of action.
- Special precautions should be taken to preserve the digital evidence with regards to its fragile nature.
- Different types of cases and media may require different methods of examination and the persons conducting the investigation should have the relevant training.
- Documentation should be complete and is an ongoing process throughout the entire examination.

NIJ empathize training and the right competence for the task, planning for the relevant scope, evidence integrity, dynamic approach to methods, and thorough documentation.

ACPO's guidelines, which are primarily written for the UK law enforcement have defined four principles (54, p. 6) for digital evidence:

1. No action taken by law enforcement agencies, persons employed within those agencies or their agents should change data which may subsequently be relied upon in court.
2. In circumstances where a person finds it necessary to access original data, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
3. An audit trail or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
4. The person in charge of the investigation has overall responsibility for ensuring that the law and these principles are adhered to.

ACPO does also focus on evidence integrity, competence and documentation. They point out the importance of a good chain of custody, good leadership and management.

A proper chain of custody requires three types of testimony (48); that a piece of evidence is what it purports to be, that the individual who had possession of the evidence from the time it was seized until it is presented in court is continually known, and that the evidence remained in the same condition the whole time.

> A piece of evidence is only as good as the Chain of Custody accompanying it. An item could be seized that has great evidentiary value, but unless the manner it was secured and accounted for can be articulated, it may be worthless in legal or administrative proceedings. (48, p. 325)

Beebe and Clark (44) argue that certain principles apply to all phases of the digital investigation and that these represents overarching goals of the investigation. These principles are *evidence preservation* and *documentation*.

Evidence preservation is based on the goal of maximizing evidence availability and quality, while maintaining the integrity of the evidence during the digital investigation process. It includes; validation from the initial incident response phase, data collection in a forensically sound matter, making forensic working copies, awareness of what steps could modify data during analysis, communication of findings in a way that facilitates future corroboration, and closure where related information from the process is retained (44).

Documentation should record all information relevant to the investigation process, for example; describing the state of a system when data is collected, what tools and methods were used. Likewise, it should include all findings – both relevant, and findings later deemed irrelevant (44).

According to Casey (36), verification of the accuracy and completeness of results is needed in each phase of an investigation. Further, an effective case management is important to ensure that the digital investigation runs smoothly and that relevant information from each step of the process is captured, documented and woven together to create a complete picture of the incident. The approaches to verification include hash comparison, dual tool verification, checking data at a low level, and peer review.

The Scientific Working Group on Digital Evidence (SWGDE) was formed in 1998 and one of its goal is to ensure the quality and consistence within the forensic community. They have published several best practice documents and guidelines. One of its focus areas is to form a basic strategy to develop confidence in the digital and multimedia evidence forensic results by identifying likely sources of error and mitigate them (55). They list the primary types of errors found in tools related to digital and multimedia evidence: Incompleteness, inaccuracy and misinterpretation. The suggested strategy to mitigate these kinds of error are to do thorough tool testing and have sound quality control procedures. This includes; personnel training, use of written procedures, documentation of examinations, laboratory accreditation, peer review, good management, and to use multiple tools and methods to complement capabilities.

By looking at data through layers we will have a better understanding of the potential for errors, and how this can be mitigated.

2.3 Digital investigation challenges

2.3.1 Layers of abstraction

There are two challenges for digital forensics with how data is stored at the lowest level; it is very complex (complexity problem), because raw data is often too difficult for humans to understand, and the amount of data (quantity problem) to analyse can be very large (56). To make data more understandable and comprehensible the process of translating data to higher levels of abstraction is introduced.

Data at the lowest level are represented in bits that are either set to one or zero, the binary number system. One abstraction example is when a text file is saved on the computer storage device; the letters are saved on the media as bits, then each of these letters have a numerical representation (ASCII), which again represents a corresponding character in the text editor. This process of abstraction translation is done by the Operating System (OS) and similarly when using a forensic tool to recover a deleted file from a storage media. This involves several layers of abstraction from the magnetic fields on the disk to the letters and numbers that we see on the screen. We just see a representation of the data, not the actual data itself (36).

Moving through each of these layers can introduce errors, so the usage of a forensic tool has the potential to introduce errors, such as incorrect or incomplete reconstruction of file systems and other data structures (36, 56). This was shown by Ayers, that even one of the most popular and well-tested tool Encase was not immune to such errors (21).

In a study by Friheim (57), investigators working with digital forensics were asked if they had discovered errors in any of the tools they had used for forensic purposes. 73% said they had discovered errors. Most of them did so by simply looking at the output that something was wrong, but many also discovered errors when utilizing a different tool.

Evidence collected from computer networks have an inherent uncertainty with potential for error and loss. Networks also have multiple layers of abstraction that hides the complexity of lower layers, with each layer providing a new opportunity for errors and losses (58).

Most tools will present data at a high level, such that it is easier for the user to read and understand. This also includes solutions based on DFaaS that uses a high-level, agnostic representations of forensic artefacts (e-mail, document, picture, files etc.) (59).

The digital investigator will require competence and expertise to be able to understand data at the different abstraction layers, both to identify evidence, but also to be able to verify the translation of data from low levels to higher levels. One practical solution to this in addition to the manual labour of verifying data at lower levels, is to use multiple tools, which is referred to in literature as dual tool verification. Dual tool verification means that the digital forensics examiner verifies the findings with a different tool. This gives more credibility and reliability to the findings (9, 48), and it enables the investigator to check the veracity of the data (36). Veracity is a characteristic of data in data science that also applies to big data, and the definition by NIST is as follows:

> Veracity refers to the accuracy of the data, and relates to the vernacular garbage-in, garbage-out description for data quality issues in existence for a long time. If the analytics are causal, then the quality of every data element is very important. If the analytics are correlations or trending over massive volume datasets, then individual bad elements could be lost in the overall counts and the trend would still be accurate. Data quality concerns, for the most part, are still vitally important for Big Data analytics. This concept is not new to Big Data, but remains important (60, p. 26).

How should one know if the tool is reliable and free of potential errors? One approach is to use open source software where the source code is available to review for the tool testers, such as the computer forensics tool testing program at NIST. This will provide for a better understanding and increases the chances that bugs are found. But since it is acknowledged that commercial tool developers will want to keep some portions of their programs private to protect their competitive advantage, it is not always possible to review the source code. One of the most effective ways of reducing potential errors and validating results are therefore through peer review. This means having another peer, such as another digital investigator double check findings and using multiple tools to ensure the results are reliable and repeatable (36).

2.3.2 Layers of trust

Using different tools to conduct our examination requires us to put some trust into the tools to give us reliable and valid results. The same counts in a situation where another person, such as a colleague or an external consultant, is acquiring a device for the investigation team. Trust is then implicitly that the person has the right knowledge, competence, and follows a forensically sound procedure.

Further, trust is put into the write blocker hardware used, they must trust the hardware interface of the forensic workstation to read the disk correctly, the integrity of the host OS must be trusted, the software used to read and acquire the disk, and the destination for the disk image that can reside on the network. These are different layers of trust that cumulates the longer away we are from the data itself (61).

Cloud environments introduces even more layers of trust. Services can be hosted externally, where trust is put into the provider of this service. In a Software-as-a-Service model the user does not manage or control the underlying cloud infrastructure. Trust will in this instance be given to the different layers of the service such as the guest OS, the hypervisor, the host operating system, the hardware and the network itself.

If law enforcement agencies wishes to remotely acquire data that is stored in the cloud, the examiner and the court must trust the integrity of the technician at the provider to execute the search in a trustworthy manner, the technician's hardware and software used to collect the data, and the cloud infrastructure (at least network and hardware) to retrieve, reassemble, and report the data (61).

Dykstra and Sherman (61) advices the digital investigator to examine evidence at multiple layers. This makes it possible to control the data for inconsistency and to correlate the evidence.

2.3.3 Why automation can be a paradox

The definition of automation according to Techopedia (62) is:

> Automation is the creation of technology and its application in order to control and monitor the production and delivery of various goods and services. It performs tasks that were previously performed by humans. Automation is being used in a number of areas such as manufacturing, transport, utilities, defense, facilities, operations and lately, information technology.

In industries, automation will greatly improve productivity, save time and cut costs (62). This is also true for the digital forensics field where it can greatly increase the efficiency of the investigation. As it would enable less skilled technicians to perform tasks according to pre-programmed parameters, thus reducing the load on experienced forensic examiners (22). Methods of automation is crucial according to Garfinkel (1), in that the only way to cope with the challenges posed by the increasing diversity and size of forensic collections is to create more powerful abstractions and easier manipulation of data. Advanced systems should be able to handle information much the same was analysts do today.

Deloitte suggests in their report on the future of policing (63) that the police should; "automate where possible, starting with areas of labour-intensive back-office processing that are relatively uncontroversial". This includes arduous manual tasks, such as hash matching (25) which would be very time consuming to do manually and would be vulnerable to errors. This is automation on a lower level which has been a standard in digital forensics for many years. But the need to automate on a higher level is ever more needed and can already be achieved with evidence collection, processing, and to some extent, documentation. Growing research is being completed, attempting to automate analysis (25). We are not quite there yet, current analytical approaches largely rely on literal string searches, simple pattern matching, indexing data to speed up searching and matching, hash analysis and logical level file reviews. These approaches lead to high information retrieval overhead and underutilization of available computational power (64).

Using more intelligent automation on a higher level can encode the knowledge of trained investigators in a repeatable, verifiable way (25). This bank of knowledge would be available for the less technical trained investigator and it would involve them more into the digital investigation. The department would have a workload reduction by filtering out unimportant devices early. This is done via triage, which is a preliminary examination that is often being done in the field by non-experts. The reduction in workload by using automation would make more time to education and deeper analysis, making experts use their depth knowledge more often, further making them more knowledgeable and competent (25).

“The forensic examination process is generally more susceptible to computer automation than forensic analysis as the latter requires some degree of critical thinking and implementation of the scientific method” Casey argues (36, p. 39). Automation can be of great help in the analysis by finding links and patterns in data that a human analyst might otherwise overlook, but the human interaction is important to interpret meaning and significance of findings (36).

Using automation on a higher level introduces new challenges. The fact that non-expert investigators with less knowledge are more involved in the digital investigation and can produce reports from the automated tools is positive. But when investigators begin to lose understanding of the underlying concepts of the investigation because of an over reliance on automation this can turn out bad. When called to giving testimonial in court on their findings they cannot establish that they have complete knowledge of how they arrived at their conclusions which could introduce doubt and reduce the credibility of evidence derived using automated processes (25, 26).

Further, the tool vendors have a monetary interest in selling forensics tools that can be used by people without the forensics experience and certifications. It has gone from a relatively small niche to a considerable expansion in the market when not just the experts can use the tools, but also case detectives and lawyers have the opportunity to apply the software (65).

Casey (66) argues that a “common mistake made by inexperienced individuals is over reliance on user-friendly or automated forensic software” and that wrong configurations (such as failing to set the correct time zone) can have major ramifications. Software bugs, which are quite common, will not be easily discovered without a critical review of the results from the tool and by validating the results in multiple tools. This can lead to misinterpretation of data or that the investigator completely misses digital evidence.

There are also arguments that claims that automation deteriorates the expert knowledge of the digital investigator. By simply pushing a button the digital investigator will soon forget how to manually conduct the same procedure because of the absence of repetition. This can in turn make them less detail oriented. Some critics also believe that the further you get away from the manual manipulation of data; the more likely there is the chance for error or omission of key evidence in a case (25, 26).

As introduced earlier, the paradox of automation is defined as “the more efficient the automated system, the more crucial the human contribution of the operators. Humans are less involved, but their involvement becomes more critical” (7). James and Gladyshev in their article (25) point out that it is crucial to implement automation correctly at the correct phase of the investigation.

With regards to implementation in the development of future tools for digital forensics, Ayers (21) lists several metrics for measuring the efficiency and performance in these tools:

- The absolute and relative speed must be improved
- Greater reliability and accuracy are needed
- Improved completeness, auditability and repeatability
- Improved human comprehension and productivity by presenting data at higher levels.

Homem (52) also lists several requirements to an automated and independent digital forensics system:

- Distribution - must handle multiple users and multiple sources of data
- Universality - must handle multiple formats and different platforms
- Responsiveness - live and remote interactivity
- Integrity - of data and process
- Privacy - protecting sensitive information
- Security - maintaining authentication, authorization, accountability and non-repudiation.

2.4 Current state of the art

To deal with the challenges that comes with big data and growing backlogs, multiple solutions have been proposed. Examples of technical solutions are: *Triage* (22, 67-72), *remote evidence acquisition* (73), *data reduction* (3, 6, 31, 64, 74-78), *data mining* (3, 16, 79-81), *distributed processing/computing* (4, 6, 59, 82-86), *cross-drive analysis* and *correlation frameworks* (6, 31, 75, 78, 87-91), *computational forensics* such as *machine learning* and *intelligent analytical techniques (intelligent analysis)* (5, 13, 37, 64, 92-95). In addition, there are multiple non-technical solutions such as legal solutions and human resource solutions (training of personnel and distribution of work among teams) (52).

Triage is a preliminary survey that involves the rapid review of many potential sources of digital evidence, with the goal to quickly identify those items that contain relevant evidence (22). This process can exclude evidence of little or no importance, such that the amount of data prepared for examination and analysis is being concentrated on the most relevant devices.

Remote evidence acquisition was presented as a method to remotely transfer an image from any suspect computer directly to a forensic laboratory for analysis (73).

Data reduction are techniques to reduce the data required to be analysed. One way to do this is by collecting sub-sets of data which contain potentially relevant data as opposed to do a full forensic image (3).

Data mining is the process of extracting useful information from large datasets, which has the potential to reduce the processing time and to improve the information quality (3). Data mining uses a combination of techniques to find this relevant and useful data, such as; artificial intelligence, statistical modelling, machine learning, pattern recognition, data visualization and database processes (96).

Distributed processing is when we move from the use of single workstations to a distributed resources of a pool of computers systems which can speed up the analysis by overcoming the input/output (I/O) bottlenecks (82, 84).

Cross-drive analysis is an approach "designed to allow an investigator to simultaneously consider information from across a corpus of many data sources, such as disk drives or solid-state storage devices" (75, p. 1). It involves the use of statistical techniques for the correlation of data on multiple disk images and it can identify which image has the most relevant information for the investigation (96). This technique together with cluster analysis, data visualization and outlier analysis reduces the data retrieval overhead, which can improve the examination time (1, 96, 97).

Correlation frameworks was termed by Raghavan (98, 99) and it describes a need for solutions that integrates multiple sources of digital evidence and identifies metadata based associations. Metadata provides information that is particularly relevant with regards to event reconstruction, which is based on situational information that can be used to determine under what contexts events transpired (98).

Computational forensics has a wide range of applications to digital forensics. Methods such as *machine learning* can be used to; find anomalies in data (e.g., malware), find material that depicts sexual abuse, implement pattern recognition, or to classify huge amounts of unstructured data (13).

Intelligent analysis combines methods from machine learning, computational modelling and social network analysis to reduce the amount of time involved in analysis (96).

All these solutions have different applications that contribute to the digital forensics process in some way or another. If using the process models depicted in Section 2.2.1, these measures can be applied to a corresponding phase. Triage, remote evidence acquisition, and data reduction would mainly apply to the preservation and collection phase. Data mining, distributed processing, cross-drive analysis and intelligent analytical techniques applies mainly to the examination and analysis phases. Computational forensics applies to multiple of the solutions and phases (96).

Digital Forensics as a Service is based on distributed processing and can include methods such as; integration and correlation of data (and metadata) from multiple sources, data reduction, and it can be expanded to include a plethora of tools that uses computational methods (52, 59).

2.4.1 Digital Forensics as a Service

Digital Forensics as a Service has been described as a cloud computing service where the computing power comes from distributed computing (14), and as a computer forensic workflow management and processing service using cloud (100). Cloud computing is defined by NIST (101) as:

> Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

The cloud computing services are split into categories such as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) (101). These define the different levels of how much of the data infrastructure that is hosted in the cloud, and what the user has access to and control over. In SaaS the end-user accesses an application (e.g., web-based email, virtual desktop, communication etc.) and does not manage or control the underlying cloud infrastructure, and the responsibility of the user is quite limited. In PaaS (e.g., runtime, database, web server etc.) the user has more control and can develop and deploy their own applications. They can use acquired applications on a hosted infrastructure which consists of the hardware (servers, processing, storage and other fundamental computing resources) and some of the software, such as the operating system. IaaS (e.g., virtual machines, servers, storage etc.) gives the user even more control and lets them run their own applications and operating systems on a hosted infrastructure consisting of hardware.

The term Forensics-as-a-Service was used to describe a concept that could "(...)make use of massive computing power to facilitate cyber criminal investigations on all levels" (102, p.14).

Lee and Hong used the term forensic cloud (103) to describe a client-server framework that would enable the forensic examiner to concentrate on the investigational process by separating the technology from investigation when providing forensic tool operations/-techniques as services.

Ayers (21) pointed out the need for new tools referring to the first generation of tools having bottlenecks in processing and I/O speed amongst other challenges. The new tools would need to handle parallel processing, better storage and bandwidth. In addition, they would need to be accurate, reliable, and provide for auditability, repeatability and high-abstraction of common data files. The forensic cloud system would provide for a scalable architecture with increased performance and parallel computing on a distributed environment, which would also give opportunities for better collaboration and better understanding of data by using data abstraction. In so it was supposed to improve the productivity and efficiency in digital forensic investigations (103).

An early prototype implementation to using distributed digital forensics was introduced by Roussev and Richard III (82). This prototype was running on a cluster consisting of several nodes, which was essentially a collection of stripped-down Linux computers that were running in parallel. This cluster was connected to a file server via a switch. The authors discovered that by using distributed processing power via clusters would efficiently speed up typical forensic functions. Pre-processing of video and images would be reduced from hours to minutes and conducting regular expression searches could be performed 18-89 times faster with only eight machines working together.

Since one of the main methods to digital forensics is searching for information, further research by Lee (104) looked at the method of indexed searching applicated to the DFaaS model. The framework that was used, Apache Hadoop, supported distributed computing that could be scaled to thousands of machines that each offered local computation and storage. The results were promising and showed an increase in the speed of indexing data for searching in large quantity of data.

Hadoop is an open source Java implementation of the MapReduce model which is a distributed programming paradigm, developed by Google (105). This model can create scalable, massively parallel applications that process terabytes of data using commodity clusters. Programs written for this model can be executed in parallel on clusters of varying size (106)

A framework called "Forensics as a Service" (100) was developed that used cloud platform to conduct forensic examination and analysis in a forensic workflow management and processing using cloud. Because of the huge potential for storing and processing in the cloud the developers demonstrated that it could save up to 87% of the analysis time compared to when using traditional methods.

In another forensics as a service model called "FRAAS" (107), the author introduced a model that was focusing on creating a repeatable system, which is one of the challenges with any cloud-computing ecosystem. This is especially true if the data is gathered externally from public cloud providers, or if data is stored externally. Although it could reduce the cost dramatically and significantly reduce the data sets analysed through the use of public clouds, which was compared to using private clouds in (108). Here the authors discussed security issues by using public clouds which is tangible/has known risks (access, availability, infrastructure and integrity) and intangible/has unknown risks. Private clouds have higher security assurance, giving the user more control and is operated by one organization on-premise or off-premise.

Security requirements for a forensics as a service model was discussed further by Marturana et al. (109). They proposed that the system should aim to guarantee security aspects of access control. It should have confidentiality and integrity of data at rest and in transit. It should promote non-repudiation and availability, as well as implement secure logging of a cloud provider's operations and data delivery to law enforcement agencies. More technical challenges in each phase of the digital forensics process with regards to cloud was identified along with recommended solutions for each challenge in (110), but it will not be discussed further, since it is outside the scope of the study.

Some of the advantages with a DFaaS model are many (14). It could speed up evidence collection and analysis. Wen et al. (100) proved that it can save up to 87% of analysis time in tested scenarios, in comparison with traditional methods. It can enable case detectives to directly connect and perform investigations themselves without waiting for expert personnel. The use of cloud computing could influence future developments of digital forensic science and open up new possibilities for collaborative investigation. By storing the data in a cloud system, it could make more intelligent processing possible, and new tools and techniques could also be integrated to further expedite the process. Instead of bringing the data to the processing tools, the tools are brought to the data. So as soon as the data is available, tools can be applied to it and the results are stored in an database that can be queried by the investigator while other traces are still being extracted (31).

Additional advantages include the low requirements for the client, where the investigator could access the system via a browser with no limitation to the position of the client. So the investigator can access the system practically anywhere (111).

2.4.2 Digital Forensics as a Service in the Netherlands

One of the more tangible examples of a DFaaS model that has been used in law enforcement over several years (since 2010), is found in the Netherlands (6). The model is implemented via a tool called *Hansken* that has been run in over 600 cases with over 2500 investigators having access to it (per 2016) (112). It started in 2006 as a scientific research project aimed at identifying and developing techniques for automating parts of the data analysis process, and has grown to be a service-based approach for processing and investigating high volumes of seized digital material (31).

In this DFaaS model the digital devices are first acquired as forensic copies, and the data is then copied to a central storage where it is examined and processed by system operators. This releases the digital investigator from many of the administrative tasks, since this is built into the model and handled by system operators. The data is sent to a centralized system that automatically extracts traces from the data and gives digital investigators, detectives and analysts access to the traces.

Detectives can log on using a web browser where the data processed from Hansken can be queried and investigated directly. Digital investigators can use the programming interface to run automated tools and analysts may want to retrieve all information and analyse the results using data visualization tools, integrate data sources, or build a network of contacts. An example of running an automated tool can be to automate some of the analysis, for example, linking the text stream of a Word document to the plain text stream of a PST e-mail message (31). This is possible due to storing each trace as multiple data streams in different formats.

The experts can concentrate more on advanced tasks such as data extraction and deeper analysis, and the detectives that is most familiar with the case, gets early access to the data and can search for evidence in the initial phase of the investigation. The responsibility and cooperation is easier when everyone can work on the case at the same time (6, 31).

The risks of running the system in the cloud made the developers focus on security above all, by making the data secure and only available for the investigators. This is done by encrypting all data and the need for an encryption key to gain access, which are stored separate from the data itself. The privacy aspect must be handled correctly, and transparency of forensic data means that at any time, it must be clear where traces extracted from the seized material originate from. This keeps the chain of custody in place. Every action of the users are also logged, which promotes transparency to other parties such as lawyers, the court etc. (31).

Using automation in the extraction of artifacts removes a lot of the manual labour from the digital investigators. It also removes the requirement for the detectives to know details about where these artifacts may be stored on the device. This removes the bottleneck of manual procedures and allows non-specialist investigators to leverage their tactical knowledge and allows technical experts to leverage their technical knowledge. The authors claim that these benefits of non-experts investigating digital material outweigh the risks of misinterpretation (31, 59).

In summary, while the amount of data grows rapidly with its diverse nature, law enforcement agencies are experiencing an increase in the backlogs of cases. This is shown in statistics from multiple countries (15, 28, 29, 32-35). The digital investigation is not something peculiar in that it differs from a traditional analogue investigation, the goals are the same, but it provides for new opportunities the police have in this digital day of age. These opportunities are regulated by processes and principles and to get along with modern demands such as big data the need to think creative and use modern methods is apparent. However, the implementation of such methods put an even greater demand on the quality assurance of the field. Several methods are listed here, and frameworks such as DFaaS combines multiple techniques to make the field more effective. If not a blueprint of how the law enforcement can cope with challenges such as big data, at least it mirrors a step in the right direction.

3 Methodology

The general approach in this study was to gather data and use it to do a socio-technical analysis (described in Section 3.1). This is a top-down approach that starts with collecting information with a focus on the users' experiences with the DFaaS model. This information was used to focus in on the problem at hand which is the quality and efficiency question of the digital investigation in DFaaS.

In the following chapter, the methods used for collecting data is discussed along with the research design and the theoretical frameworks that was used to interpret this data.

Section 3.1 lays the foundation of the theoretical frameworks used in the thesis. These are the methods that are used to interpret and analyse the data. **Section 3.2** describes the actual data that are collected and puts them into context by applying it to the frameworks. **Section 3.3** explains the methods that was used to gather data in the research, both in theory and practise. The validity and ethical concerns by using these methods are also discussed. A summary follows last in the chapter.

3.1 Theoretical framework

To interpret, correlate and systematize the data in the research, two theoretical frameworks were used. The socio-technical systems model was used as an overarching framework to categorize the different parts of the DFaaS model in an investigation. And the digital investigation process model was used as a guide of the investigation process, with a focus on how the concrete tasks in an investigation plays out in the DFaaS model.

The socio-technical systems model

The socio-technical systems model has had a long history. It was developed in the middle of the 20th century when technology started to impact business efficiency and productivity. In this regard there were many examples of technology being associated with implementation problems. Researchers suggested that a fit between technical sub-systems and the social sub-systems was needed. The model was updated later to encompass the introduction of ICT (113).

Kowalski (8) used the term information-technology (IT) insecurity as an expression to describe vulnerabilities in an IT system. One way such a system could be vulnerable is when the system is not in balance or when it has not reached homeostasis[6].

A change in the machines used in the system can not only affect the methods used in the system, but also the structure and the culture. One example of this is if a corporation were to install an advanced and modern firewall to protect the business, one risk could be that the IT personnel either did not have the competence to handle security warnings or they put too much trust in the device. The consequence could be to ignore real threats and focusing too much on false positives, or that the firewall is not automated enough and

[6] Homeostasis in this setting equals a balanced system and is the underlying principle (8) that describes an ideal situation where the different parts of the system (the social and technical) are making each other better, and finally the whole system better.

requires manual intervention to capture intrusion attempts. Or maybe the firewall was not adapted properly to the structure and did not take into consideration third party vendors with VPN access to the corporation. To have good IT security, the system will need to be in balance.

The focus of using this model in an IT system is often to mitigate security issues, but one could argue that it could be used to measure quality as well. Quality in this regard relates directly to the overarching principles in an investigation, such as evidence integrity and chain of custody, which was listed in Section 2.2.2. Without it, the rule of law will be jeopardized. It can then be argued that quality provides some kind of security in terms of legal security.

This socio-technical system model will be used as a framework to understand the inner workings of the DFaaS model and how this relates to the digital investigation. The socio-technical model gives a very broad and holistic view since it takes every aspect of a system from the machines and the methods used, to the culture and the structure.

Figure 9 illustrates the two main parts of a socio-technical system; the technical and the social. The technical part is further divided into methods and machines, while the social part is divided in culture and structure. The arrows illustrate the interrelationship between the parts, as a change in one part will affect the other parts and the system as a whole. That is why the weight balance in the middle illustrates the importance of having the right balance in the system for the system to have homeostasis[6] and thus promote both, in this case, efficiency and quality. If there are too much weight on one side, as illustrated in the figure, the integrity of the system will be compromised which could for example lower the quality output of the system.

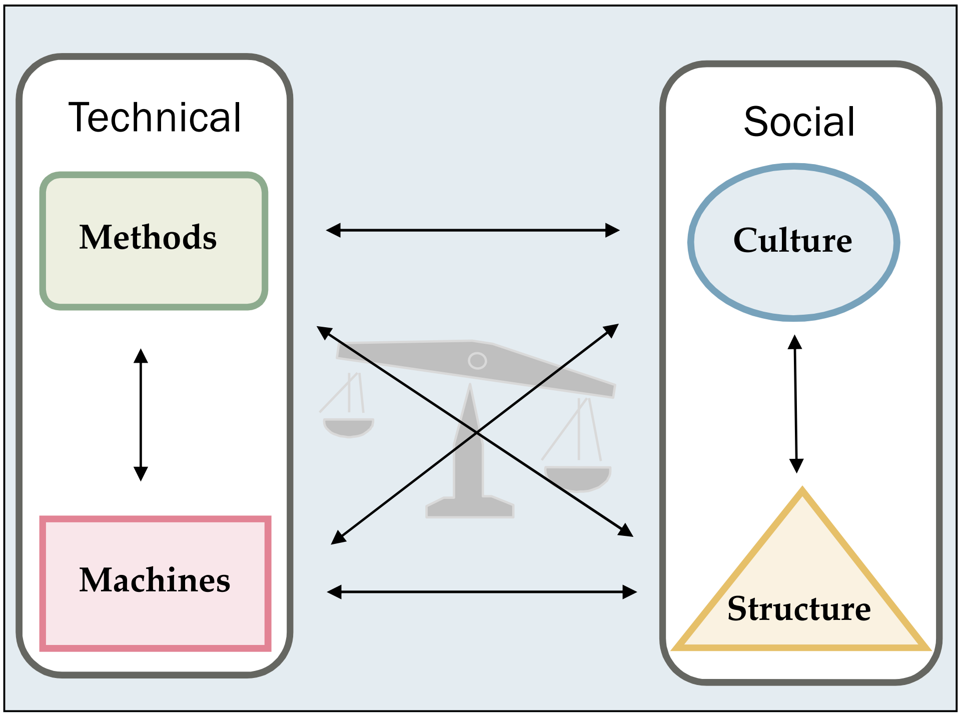


Figure 9: A socio-technical system (based on Kowalski (8), page 10).

Hybrid digital investigation process model

In Chapter 2, the reason for using process models in a digital investigation was identified, such as promoting the quality and standardisation of the investigation process. The chapter also looked at the evolution in the developments of these process models. Further, several principles for the field of digital forensics was identified to reach a high standard with the rule of law being the utmost priority.

In this thesis, a process model was used as a theoretical framework for the methods used in a digital investigation. Since there are a plethora of process models that have been developed (10, 14, 36, 39, 40, 42-44), some requirements were set when trying to identify a model that could fit into a socio-technical analysis of DFaaS:

1. The model should be detailed and include all the steps/methods involved in a digital investigation, and therefore have a high degree of relevance.
2. The model should be easy to understand with clear definitions.
3. The model should have some underlying concepts of quality management.

These requirements were identified based on the task that would enable work methods in a digital investigation to be measured against the application in a DFaaS model. To measure whether an underlying quality management supports this model, this should be an integrated part of the process model itself.

Different process models were reviewed earlier in the thesis (10, 36, 42), and none fulfilled the requirements completely. Because of this, a hybrid of multiple models was created by the author, as seen in Figure 10.

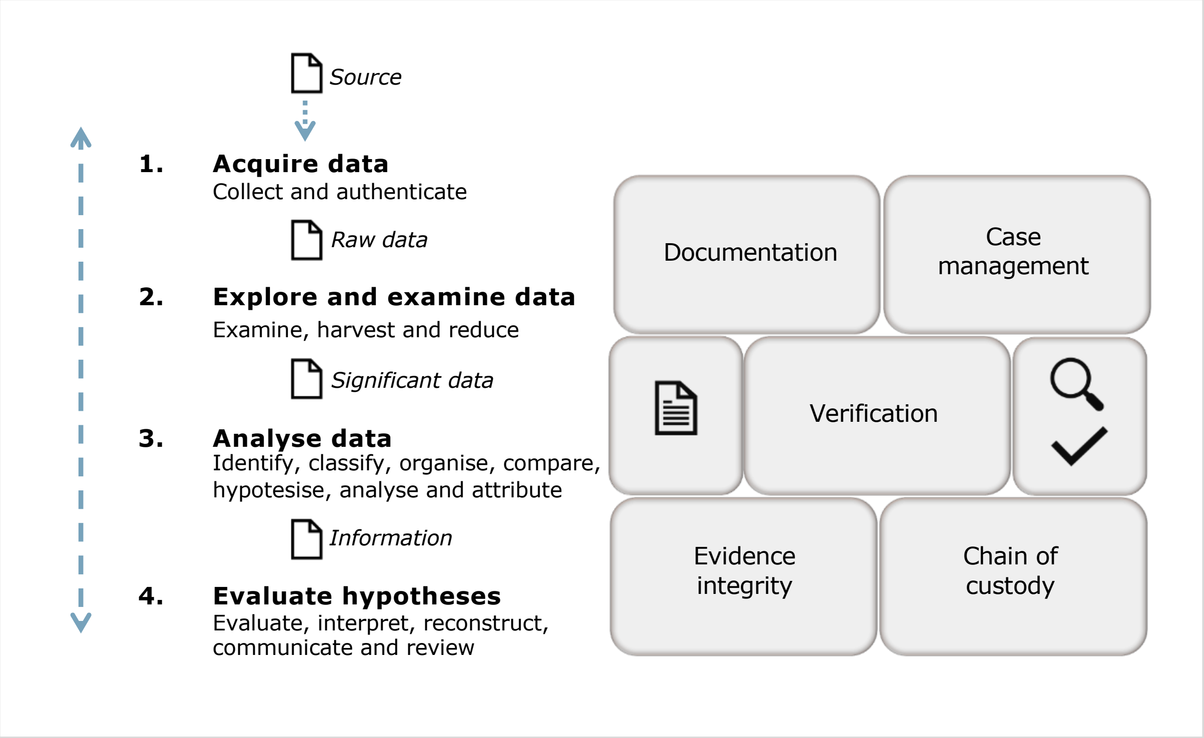


Figure 10: Hybrid digital investigation process model

The model is given a more detailed explanation in Figure 11. The hybrid process model includes the *methods*, the associated *tasks* and the *data description* for each step. On the right side is the *scaffolding principles,* which works as overarching quality markers. These are illustrated as a brick wall to emphasise its necessity to have a good foundation for the investigation.

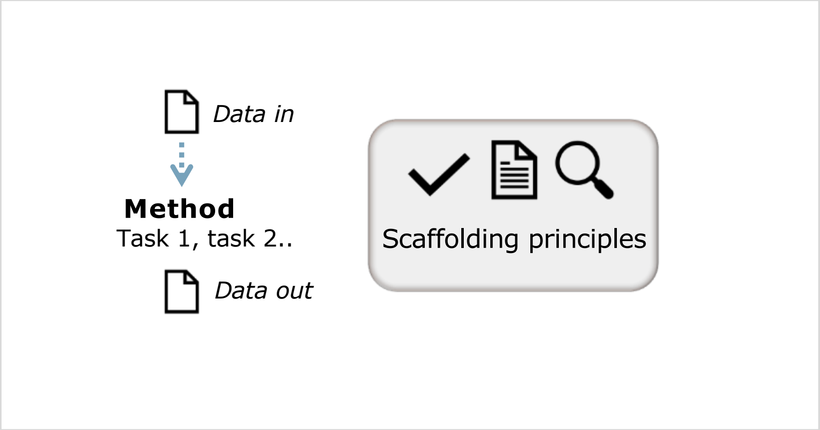


Figure 11: Explanation of the steps in the hybrid process model

Methods: To make the model easier to work with, the sub-phases were gathered under four main methods that describes the underlying tasks involved, these were: A*cquire data, explore and examine data, analyse data* and e*valuate hypotheses*. The three first methods are taken from the Data collection and processing process model (10) described in Section 2.2.1. The last method is from the same article, but at the investigation level in the Criminal investigation process model, described in Section 2.2.

Data description: The data description or data in/data out, describes the abstraction level of the data from source, to raw data, to significant data, and to information. These are taken from the same models mentioned in the previous paragraph.

Tasks: The Integrated digital forensics process model (IDFPM) (42) was used as a divided baseline for the tasks involved in a digital investigation. The reason being that it is significantly detailed with concrete tasks. The model was also used as a theoretical foundation by the developers of Hansken in the Netherlands (6), therefore it should have a high degree of compatibility with the actual target group. In this regard it should be better to have too many steps, than risking missing a step when doing the examination of the work methods implemented into the investigation model (DFaaS). Using all the steps is optional in most situations but it needs to be fitted accordingly to the case type.

When doing a socio-technical analysis it is important to define the system in focus, and the systems above and below (8). The system in focus with regards to the process is the *digital forensic investigation phase*[7], the system above is the incident response phase and the system below is the presentation phase (please refer to Figure 7 on p. 15 for the complete model).

[7] In this thesis, simplified and referred to as the digital investigation phase.

Scaffolding principles: For quality control measures, the concepts *verification* and *case management* are included to show important practical measures. These are taken from "Scaffolding for digital investigations" of Casey (36):

> Verification of the accuracy and completeness of results is needed in each phase of an investigation. Effective case management is one of the most important components of scaffolding, helping digital investigators bind everything together into a strong case (36, p. 197).

These are built on principles like *evidence integrity* and *chain of custody* (9), that was reviewed in Section 2.2.2. In addition, the art of *documentation* which makes up the chain of custody was included. This was emphasised of several authors as important (9, 44, 53-55).

The two-pointed arrow to the left in the model is illustrating that the process is iterative in nature, which was emphasized in (43, 44).

The next question that naturally arises is if this model is compatible with DFaaS. This will be discussed further in Chapter 5.

3.2 Scope of data sources

In the introduction, the socio-technical systems model was presented with the following categories and subdivisions attached to each part of the model (technical/social) as presented in Section 1.4.

Technical part: A process model represents the methods and an application constitutes the machines.

Social part: The people investigating embodies the culture, the organization and the infrastructure creates the structure.

Based on this, the actual data sources that was used in the study is presented in Figure 12.

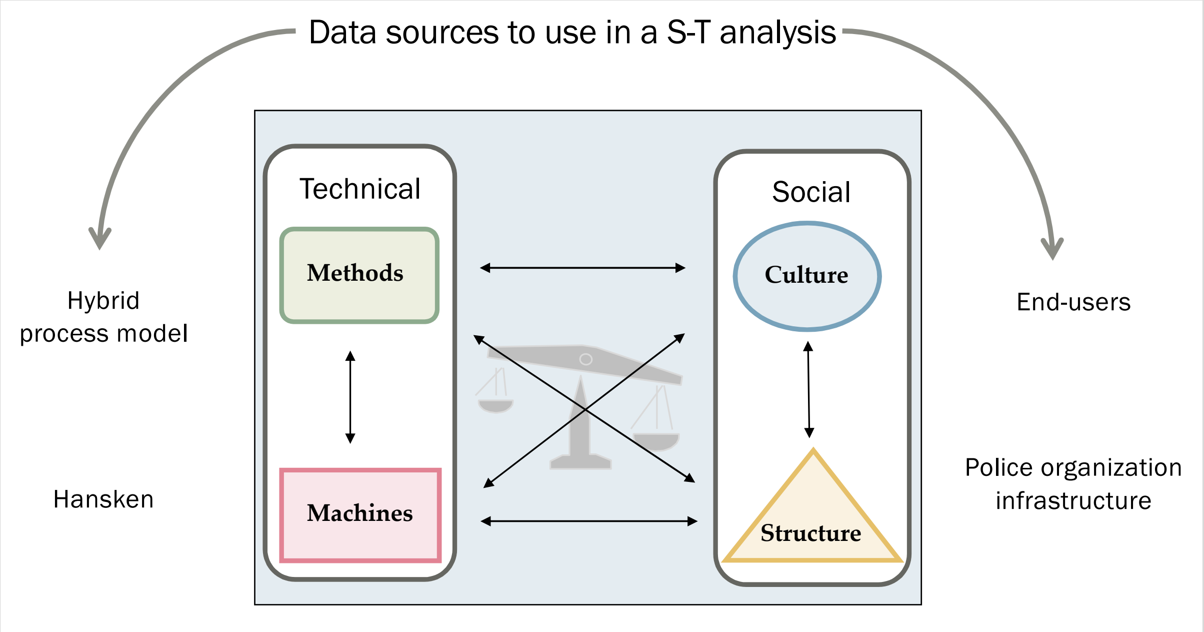


Figure 12: Socio-technical model and data sources

Hybrid process model: The process model created in the last section (see Figure 10) was used as a baseline for the methods used in the investigation.

Hansken[8]: The system/tool "Hansken", will be used as a representation of the machines. This is used in the Netherlands, which is described in (6, 31, 59), and in the second chapter of this thesis. This is just one example that have been in production for several years. The intention is not to provide a review of Hansken in particular, but to generalise the findings to DFaaS in general.

End-users (of Hansken): The end users consist of digital investigators, detectives and other parties directly tied to the usage of DFaaS. This makes up the culture. System administrators and developers will have a more peripheral role in this study.

Police organization infrastructure: Represents the structure. This could be anything from the facilities (offices, laboratory etc.) to the network hosting the DFaaS system. Also, more peripheral institutions, such as the police academy will be part of this.

On relevance

These different data sources should have a high degree of relevance and compatibility with an actual digital investigation setting. Hansken is very much an operational system, that is well documented with a transparent development.

It has been in use for several years and there should be lots experienced among the people that have used this in investigations. By reaching out to this group of people, a direct link is provided to this bank of knowledge. This is crucial to meet the main motive for this thesis, which is to answer the research question. The methods for reaching this objective is drawn up in the next section.

3.3 Choice of method

The requirements that were set for the choice of method in this thesis, was that it should be suited to answer the research question. It bears relevance to repeat this:

How can law enforcement agencies use Digital Forensics as a Service to meet the modern digitisation challenges with focus on fundamental investigation requirements?

Digital Forensics as a Service was investigated through the literature review in Chapter 2, as well as the challenges that face law enforcement agencies in our modern society. This in terms of the ability to handle huge amounts of diverse data sources and still be able to fulfil the requirements and principles defined by a high investigation standard.

[8] It may be that Hansken should have a much greater part of the system. Hansken is more than just an application, and it would be more fitting to call it a system that consists of several parts in of itself with e.g., different interfaces in the front-end, multiple service modules in the back-end, as well as logging services (31). Due to the scope of the study, and to make it possible to see it in a socio-technical context, some rendering had to be done. That is why Hansken represent the machines throughout this thesis.

Method

To answer the research question, fully or at least partly, and to increase the understanding of the topics and maybe draw some conclusions, a combination of *survey research* and *case study* was used. The reason for using survey research in this study was to have an understanding on the human relationship with DFaaS by gathering data from people with first hand experiences. Since the focus was mainly directed at a concrete model, DFaaS, and the data was enriched with comments and a structured interview, this can be defined as a case study. But the primary most tangible method used, was survey research, which ties into the research design applied in this study.

Research design

The research can be classified as either using quantitative or qualitative methodologies. Both involve similar processes, like identifying a research problem, reviewing related literature, and collecting and analysing data. But by definition, they are suitable for different types of data: Quantitative studies involve numerical data, whereas qualitative studies primarily make use of nonnumerical data (114). Both methodologies can also be combined into a mixed-methods design. This often gives a more complete understanding of the research problem, but it is also often more complex and time consuming.

For this thesis the *mixed-methods design* was chosen because it provides for a holistic view which is appropriate when using the theoretical framework, the socio-technical systems model.

Further, an *embedded design* was selected because both the quantitative and the qualitative data are collected within the same time frame, and the quantitative approach is dominating with the qualitative approach having a supplementary role. This was largely done to reduce the time requirements and the complexity of the research endeavour because of the scope of the study.

An example of using the embedded design is when planning a survey with statements which participants either "agree" or "disagree" at various points along a rating-scale continuum. At certain points there will be several open-ended items which participants could optionally explain their findings. This will give quantitative data enriched with qualitative data to help the researcher make better sense of the numerical findings (114).

3.3.1 Procedures for data collection

The scope of the data sources was explained along with the methods to use when gathering this data on a more theoretical level in the last two sections. In this section, the more practical approach on how the data collection was done will be further explained.

On the selection of the survey sample

Survey research involves acquiring information about one or more groups of people, such as their characteristics, opinions, attitudes, or previous experience. The researcher poses a series of questions to willing participants, summarizes their responses, and then draw inferences about a population from the responses of the sample (114).

When choosing the selection, the system in focus will need to be defined. There are different levels of socio-technical systems; international-, national/branch-, organization- and individual/group system level (8). Since the people involved in the investigation belong to law enforcement, the focus will be at the organizational level.

The system above on the national level, consists of the national authorities such as the government and the parliament. Actors on this level develop regulations and authorisations which defines the legal basis for the investigation.

The system below consists of people that are not directly part of the organization, for example suspects, victims or witnesses. This level is indirectly or directly affected by the system in focus such as when investigators finds evidence via the DFaaS system, this could eventually lead to a court decision for the suspect.

The following requirements was established with regards to recruiting the sample selection used in the survey:

1. The users should have experience with Hansken, or its predecessor Xiraf. It should be easy to filter out if the respondent does not have this experience.
2. The users should have variety in terms of: Background, role, experience, gender, working location, digital expertise and if they are police or civilian.
3. The minimum threshold was set to 1% of the total users, although a higher number was preferred. Based on numbers from 2016, there was about 2500 users of Hansken in the Netherlands (112). So, the threshold was 25 users.
4. The survey should be spread to at least ten-times the minimum threshold of respondents, which accounts for 250 users.

On the content of the survey

The survey (shown in Appendix C) sought feedback from the actual users of Hansken; their experience of using it in digital investigations. It focused on subjects like trust, satisfaction and how it has affected their investigative work. The goal was to investigate the user’s perception on whether they found it suitable to meet the modern digitisation challenges with focus on investigation quality and efficiency with DFaaS.

Regarding the efficiency of the system, and the compatibility of the work processes, several statements were put forth that focused on the tasks from the hybrid digital investigation process model, depicted in Figure 10.

Examples were “In Hansken we get access to the data early in the case” or “In Hansken we can strengthen/weaken hypotheses quickly”. These statements, the respondents were asked to rate according to how much they agreed to them along a rating continuum split in four (1 - “absolutely”, 2 - “to some extent”, 3 - “not so much” and 4 - “no).

The users were also asked several questions regarding the quality procedures in the digital investigation, with questions like “how much do you trust the tool to give valid data?”. They were asked questions on verification procedures, such as; if they used multiple tools to verify data, if they were verifying work of others or had their own work verified. Also, if they found it easy to validate data.

Structured interview with the Netherlands Forensic Institute (NFI)

In addition to the survey of the end-users of Hansken, some questions were developed and sent over to the NFI (see Appendix D). These questions sought to clarify some of subjects that were not specifically clarified in their paper, for instance how data is verified in the tool and updates regarding number of users etc.

Test report from the NCIS (Kripos)
A test review report (115) by the Norwegian NCIS (Kripos) was included in the analysis. This report was exempt from public disclosure, however the information permitted to include was established with the author. The review by NCIS was conducted on the behalf of an assignment given from the Norwegian Police Directorate, with the intention to look at Hansken as a possible service model in digital investigations in Norway.

3.3.2 Validity and ethical concerns

Validity
There are different strategies to support validity of findings in research. In quantitative research, two parameters are often used which is internal validity and external validity (114).

Internal validity is the degree of which the researcher can draw accurate conclusions about cause-and-effect and other relationships within the data. One example to strengthen internal validity is to use triangulation, which is using multiple sources of data to support or refute a hypothesis or theory.

External validity is the degree to which its results apply or be generalised to situations beyond the study itself. To enhance the external validity, one may use real-life settings, in contrary to laboratory studies, or representative samples.

Creswell (116) mentions multiple validation strategies such as; prolonged engagement and persistent observation, triangulation, peer review, refinement of working hypotheses, clarifying researcher bias and external audits etc. He recommends that qualitative researchers engage in at least two of them.

In this study, triangulation was applied by using multiple sources of data (survey, questionnaire, papers, articles and test-report). The research problem questions were dynamic and have been refined throughout the process. Potential personal bias has been clarified, and the supervisors were continually consulting and providing auditing on the study.

Since the study also has elements of qualitative research, parameters such as internal and external validity do not necessarily apply to the same degree (114).

The study’s external validity is discussed further in Section 5.6.

Ethical concerns
Whenever humans are involved in studies there are multiple ethical concerns for the researcher to be aware of. They can be described in four different categories: protection from harm, voluntary and informed participation, right to privacy and honesty with professional colleagues. In addition, the researcher must obtain permission from the appropriate committee at their institution to conduct the study (114).

Before the data collection was started in this study, a notification form was submitted to the Norwegian Centre for Research Data (NSD) for approval. NSD is the Data Protection Official for Research for all the Norwegian universities, university colleges and several hospitals and research institutes. The NSD concluded that the data collection methods did not contain information that could identify any individuals, and that the project could be carried out (Appendix A).

The participants of the study were not being exposed to any physical or psychological harm. The participation was strictly voluntary and based on informed consent, in that a description of the nature and goals for the study was disclosed to the participants. This included a description of what the participation would involve in terms of activity and duration (Appendix B). The confidentiality of the respondents was guaranteed, and they remained anonymous, including the author and supervisors.

In summary, as explained initially in this chapter, the general methodology in this thesis was to gather data and use this data to do a socio-technical analysis. The data basis gives a foundation on the usage of DFaaS in a digital investigation, and by using the socio-technical approach, the quality and efficiency concerns of the digital investigation can be analysed.

Two theoretical frameworks were presented; the socio-technical systems model – which has been around for years and used in a multitude of research, and a hybrid digital investigation process model was created – that merged multiple existing models based on requirements set in advance.

The methods used in the thesis was shown, which was a combination of survey research and case study of the DFaaS system Hansken. Some requirements to the selection of the survey participants was defined along with the general topics of the survey. Additional data sources were an interview with the NFI and a test-report from the Norwegian NCIS (Kripos). The use of multiple data sources should increase the validity of the study.

4 Experiment and results

In the following chapter, the two theoretical frameworks used to develop the survey; the socio-technical systems model and the hybrid digital investigation process model (explained in Section 3.1) was further used in interpreting the answers from the survey.

Since the study was based on mixed-methods, there was a combination of pure numbers (ratings and statistics) and qualitative data (comments).

The main topics for the survey was the investigation requirements efficiency and quality. **Section 4.1** shows how the sample was collected and it gives an overview of the characteristics of the users. **Section 4.2** looks at the results regarding questions on the work process used in the investigation. The answers will be measured and given a score based on whether the users find the work processes to be fulfilled its potential in the DFaaS system. **Section 4.3** addresses questions regarding verification and quality control, and whether it has improved their work in aspects such as quality and effectivity. After this follows a summary.

4.1 Sample overview

Sampling procedure
To collect the sample in the study, nonprobability sampling was chosen. With this approach there was no way in predicting or guaranteeing that each element of the population would be represented in the sample. The reason is how the users was contacted in this study, there was no feasible way to contact all users of Hansken since there was no index of users.

Recruitment of the sample
Based on the requirements set in Section 3.3.1 the author contacted NFI and asked for contact details on Hansken users. NFI gave contact info to a product owner of Hansken, with access to a database of users. The concerned was contacted and confirmed he had access to 100+ users. This was below the threshold, so the author further contacted a police liaison in the Netherlands and got contact details to an additional product owner with access to about 100 users.

These users in these databases were distributed across the country and contained users from both the regional and the national level of law enforcement. Since this was below the threshold set to 250 users, the social network LinkedIn was used to get a broader selection. Using the snowball effect, the survey was distributed to people based on their work in the Dutch police[9]. This consisted of about 70 users where some of these offered

[9] Search filter based on: Work place – police in the Netherlands, position – investigator and/or education with in cyber or forensics.

to share the link to the survey with other colleagues. An approximate of the total number of potential respondents based on this would be around 300 users[10].

The number of actual respondents was 28 people, which was above the threshold and in line with the requirements set for the sample selection.

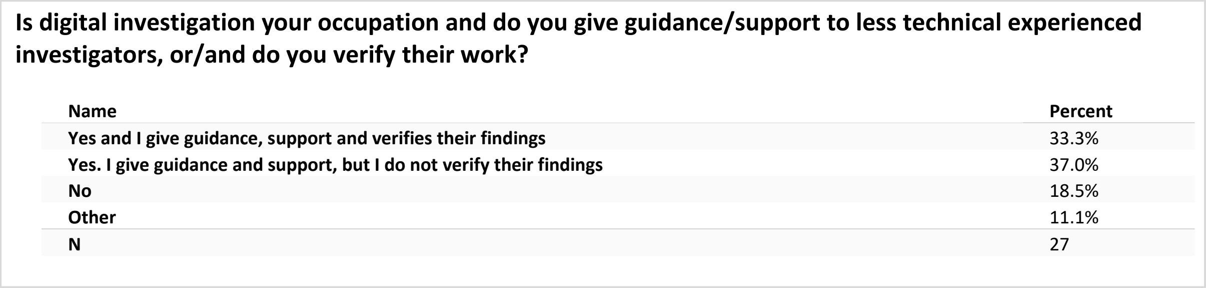
Characteristics of the users in the sample

There are different roles in a DFaaS system, hence it would be advantageous to have a diverse user group in the sample.

Their characteristics were as follows: 77.8% were male and 22.2% were female. Over half of the participants were experts in the field of digital forensics or cybercrime combined (48.1%), and 18.5% were tactical investigators. In addition, there were users in the field of education, administration, development and supervision.

70.3% had digital investigation as their main occupation and gave guidance and support to colleagues. Almost half of these verified others work as well. 18.5% did not do digital investigation as a main occupation (Table 1).

Table 1: Respondents who had digital investigation as an occupation



Is digital investigation your occupation and do you give guidance/support to less technical experienced investigators, or/and do you verify their work?

Name	Percent
Yes and I give guidance, support and verifies their findings	33.3%
Yes. I give guidance and support, but I do not verify their findings	37.0%
No	18.5%
Other	11.1%
N	27

Most were police educated (74.1%) and had more than 13 years of experience (63%). There was an almost even split between national (37%) and regional (40.7%) police unit association.

Regarding the experience with using Hansken, 43.5% had 1-2 years of experience, 17.4% had less than a year of experience, and the rest (39%) had more than three years of experience. Almost all (91.3%) had worked with digital evidence before using Hansken. 34.8% used it monthly or less, while 13% used it daily. 14.8% had never used Hansken or its predecessor Xiraf, the main reason was that it was not used at their work place. These did not participate in the rest of the survey specific to Hansken.

[10] It is hard to estimate exactly, since the users could overlap over multiple databases and the amount of forwarded invitations to the survey is unknown.

4.2 Socio-technical mapping of the work processes

In this section, the work processes of the hybrid process model; which constitutes the methods in the socio-technical system, was mapped to the users; which embodies the culture of the socio-technical system. This was done by going through the tasks of the hybrid process model and look at the feedback from the users of Hansken on their experience with this in digital investigations.

The machines and structure are also an important part of this experiment but have a more indirect role. An example is when looking at the tasks related to the analysis phase and how they are carried out by the users. All this happens on the Hansken platform, which constitute the machine, and the service is hosted on an infrastructure where the users get their access, which forms the structure part of the model.

The respondents were given multiple statements that had the following alternatives:

- Absolutely
- To some extent
- Not so much
- No
- Do not know

The ratings were 1 for "absolutely", 2 for "to some extent", 3 for "not so much" and 4 for "no". The alternative "do not know" was not included in the rating.

1. **Acquire data**

This step involves gathering the data sources and bringing it into raw data. The *collection* step involves making a working copy of the original media that was seized. The creation of an image from the original device could have been done in the previous phase (incident response), for example if the physical media is not feasible to seize, then this must necessarily be done at the scene. After the copy has being made, its validity should be verified with a hash checksum against the original to see that it is exactly the same. This *authenticates* the data (42).

In Hansken the first step is to copy the image to a central database, so the act of creating images of the devices must have been done separately beforehand. This step should easily be integrated into the DFaaS model, with the same requirements to create images in a forensically sound way using write blockers and read only software. In the test report from Kripos (115) it was reported that the bottleneck would be related to the disk imaging process, not to the processing of data in Hansken, even in a scaled down test system like Kripos used. Hansken is very adaptable and could easily be fitted with extended capabilities via customized scripts. Kripos tested this functionality and created a script that automated the process of uploading images. Utilizing opportunities like this, that would need manual handling otherwise has the potential to save a lot of time.

Regarding the quality control and verification, the data integrity and authenticity is validated by creating an index file that contains calculated hashes of individual blocks of data. Further, the data communication in the uploading process is encrypted, so is the data that is stored in Hansken with a cryptographic key that is only available to the investigator who set up the imaging process. For users to gain access there are user management in place with authentication and authorization of users, for example via Active Directory (31).

In the survey the respondents were asked whether they had access to data early in the case. 38.1% answered "to some extent" and 28.6% answered "no". The average score was 2.52 (Figure 13). One of the explanations was that the upload process was slow, and the software was unstable. Another comment to why it was not given full score was that the work process was not uniformly defined.

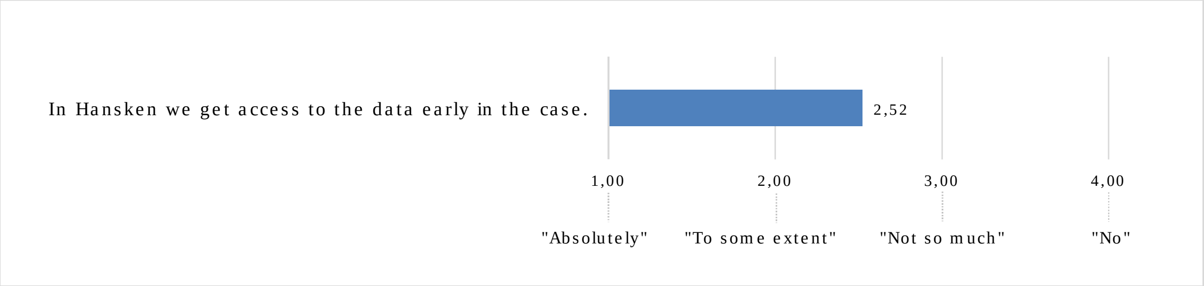


Figure 13: Acquire data survey questions - averages

2. Explore and examine data

Here the raw data is explored by the investigator and structured into significant data for the investigation. The *examination* of the data involves making the data visible and extract it into human readable form. This includes data that is hidden or obfuscated/deleted. *Harvesting* is done by giving the data a logical structure, and further taking the raw data and transforming it into information the investigator can use. By using hash sets, such as the National Software Reference Library (NSRL) the data can be *reduced* by excluding known elements (42).

In Hansken the images are processed with a standard set of tools that extracts file systems and files, carves for deleted files, or parses chat logs, internet history or mail databases (6) - to give some examples. These tools can be expanded to include custom tools, and new functionality are continuously developed. To reduce the data, file hashes can be matched against hash databases.

The processing can be automated to start as soon as the data is read from the image where it is kept in memory and multiple tools are applied to it. Instead of the tools reading data from the image, the data is already available. NFI calls this "bringing the tools to the data" (31).

The respondents were asked if Hansken gives complete data (e.g., deleted/unallocated, hidden - or obfuscated data). Nearly half (47.6%) answered "to some extent", and 28.6% answered "absolutely". This gave an average score of 2,05 (Figure 14). One comment was that it was hard to judge whether the data was complete or not, and another one found it hard to distinguish the source of the data especially when data from lots of devices was in the database.

On question whether the amount of "data noise" (such as system files) was reduced in Hansken more than half (57.1%) said "to some extent", 19% answered "not so much", and 9.5% totally agreed. The average was 2,21 (Figure 14). Some of the users with more technical experience commented it was possible to filter such files away, but it could be hard to discern what was system files for a tactical user.

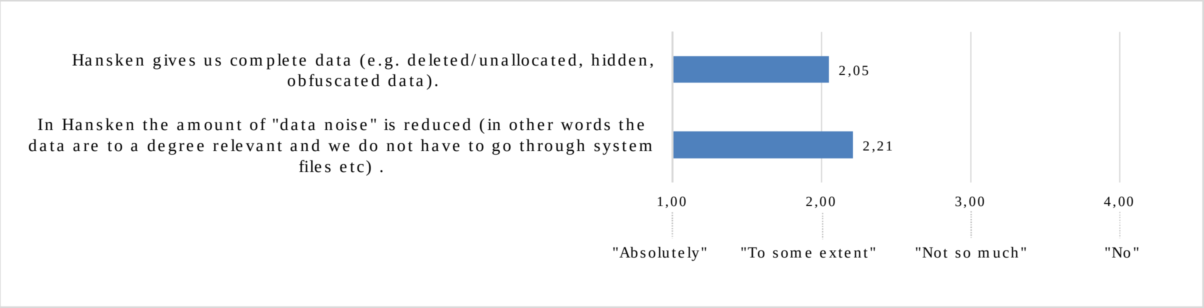


Figure 14: Explore survey questions - averages

3. Analyse data

The analysis is when actual information is extracted from the data. It depends on the case and the expertise of the investigator what this phase includes in practise. It may be only some of the sub-steps. The evaluation of information in the next phase will generally mean that an investigator with the right technical and tactical competence has to be closely involved in the investigation, since digital evidence evaluation require both (47).

The *identification* of the incident to be investigated can be known before the analysis starts in the form of *hypotheses* that has been created. But the analysis can also help identify additional incidents, with additional hypotheses to be added to the investigation. *Classification* and *organization* of the data gives structure to the digital investigation with the right focus based on the incident type and similar incidents can be *compared* with the current case. The *analysis* is conducted based on the formulated hypotheses and evidence found can be hypothesized to be *attributed* to a specific user.

Hansken makes it easy to access data either via a graphical user interface or via a programming interface depending on what kind of information one wants to access. Users can log in via a web browser and start searching for relevant data or run automated tools and query the database directly via programmatic interface (31).

Some of the benefits of using an service based model was identified in (6). There is a centralized capacity that supports the investigation by doing the administration outside of the investigation itself. This releases the digital investigator of many of the manual tasks that steals valuable time that can be used to support the case detectives with their valuable knowledge to complete more in-depth analysis. Detectives get easy access to the data and can make new hypotheses early based on their findings. All in all, Hansken seems to have the potential to strengthen the collaboration between case investigators and digital forensics experts, it gives easier access to the data and makes investigators more effective, and it puts the right competence at the right place in the investigation.

The results from the survey is depicted in Figure 15. All the responses are under the average level of 2,50 with a total average of 2,00. This means that the respondents agree more than less with regards to questions on the work processes of the analysis phase.

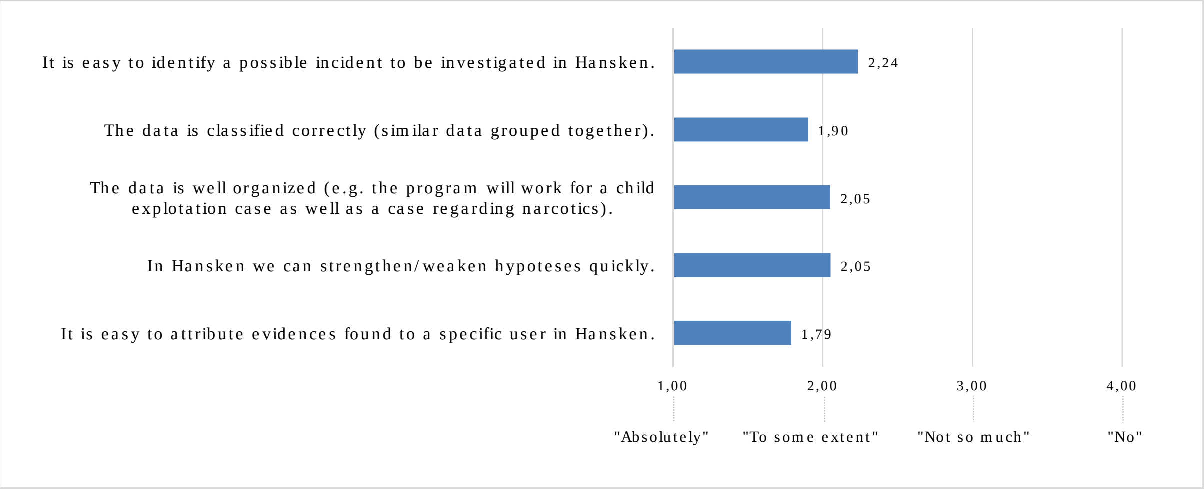


Figure 15: Analysis survey questions - averages

The users generally think that the data is well organized and that it works for multiple case types, but that there may be programs specialized for i.e. child exploitation cases that may be better for such. The fact that "all-in-one tools" cannot be great at all case types was also mentioned in the test report by Kripos (115), but it was also listed as an advantage because traces will be better merged and the investigator will avoid a disjointed review of the data.

Regarding the user interface they found it versatile, but that it is targeted more at beginners that could have problems explaining what the data actually show or the meaning of it.

On collaboration

As was pointed out in the analysis phase and in Section 2.4.1 one of the drivers of a DFaaS system should be that it fosters better collaboration. Most of the respondents think that after the introduction of Hansken they have better collaboration on cases (see Figure 16). They think collaboration and sharing of (parts of) data is much easier with Hansken.

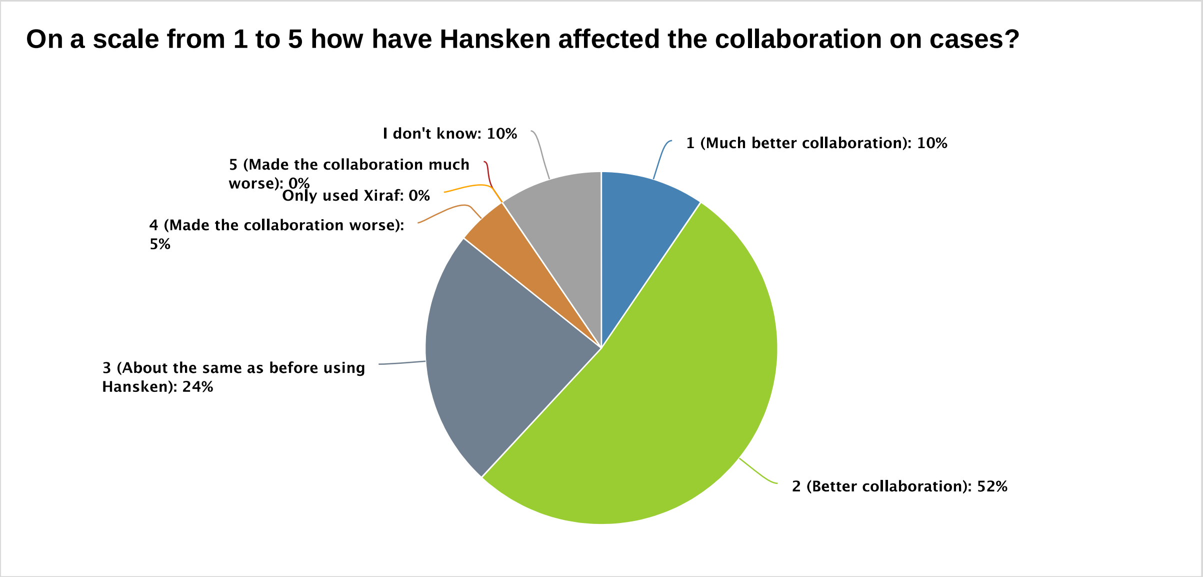


Figure 16: Collaboration on cases

4. Evaluate hypotheses

This step is closely related to the analysis since it involves the testing of hypotheses that were formed in the analysis. The information we have from the analysis is *evaluated* whether it holds true. The *interpretation* is also part of this hypotheses testing, by extracting meaning from the data. *Reconstructing* events is an important part of the hypothesis testing. We generally want to know if the hypothesized event is possible, and if the traces found can be connected to the incident. Findings from the digital investigation are *communicated* to the investigation group and the results are *reviewed* and conclusions are formed.

Questions regarding the steps of the evaluate hypotheses phase (Figure 17) had a total average of 2,07 which means that most partly agree to the validity of the methods in Hansken. The question regarding reconstruction of data was the question that had most users (28.6%) say that they did not know.

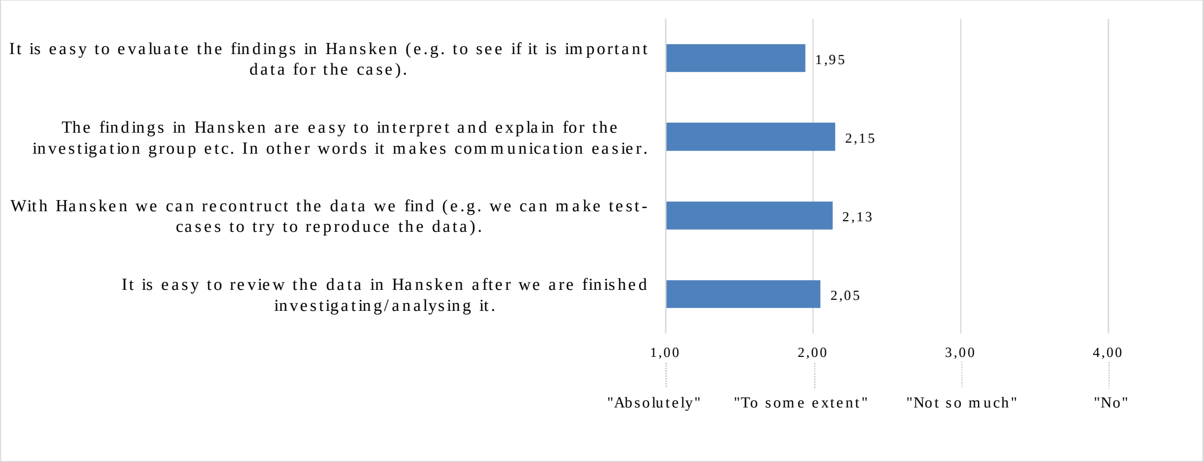


Figure 17: Evaluate hypotheses survey questions – averages

Methods regarding quality control

The hybrid model also pinpoints some general quality measures such as documentation, verification and case management. Figure 18 shows the respondents were asked whether they found Hansken to made it easier to manage a case and have a good overview of the process. Most did partly agree. One comment was that it is mainly used for storage and analysis, and that it tells nothing about the further process of a case. On question whether it was possible to verify and validate the quality and correctness of data through the investigation process, most did also partly agree to this. One commented that it was difficult to validate the uploads without a hash for the entire image.

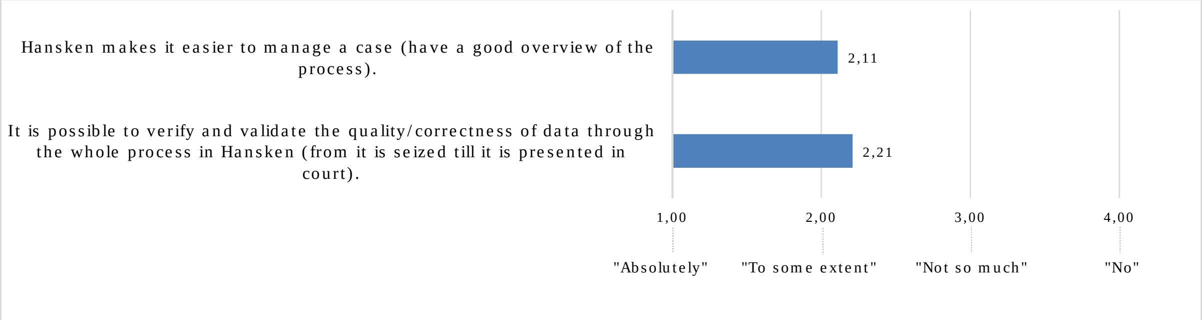


Figure 18: Quality measures survey questions – averages

According to the NFI there were checks in places to verify the image, errors would be logged and possible to investigate further. These technical validations are done by the case operators that are required to have to knowledge to do this. Further validation and verifications of results and conclusions from the digital investigation are up to the digital investigator.

Overall average score
The average score for all the questions regards methods in the DFaaS model was 2,10 – this means that most users seem to partially agree to some extent (Figure 19).

In other words, it can be argued that given the score, the users find the methods to be fulfilling their potential in the DFaaS system. But the number are too vague to give any clear conclusion. This is being investigated further in the next chapter.

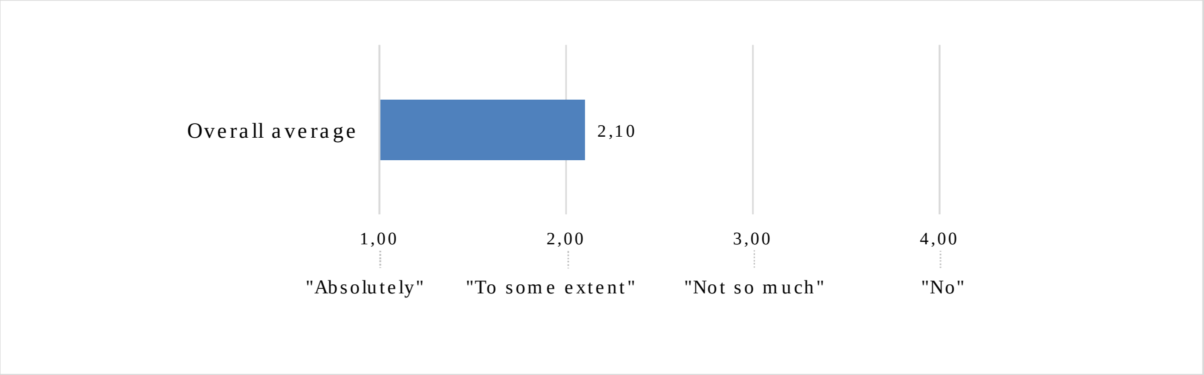


Figure 19: Methods survey questions – average score

4.3 Mapping the fundamental requirements

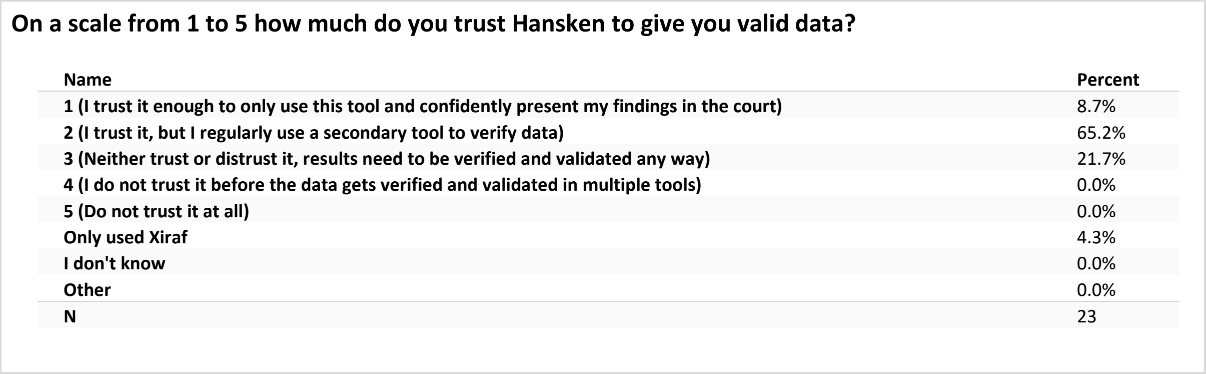
Section 2.2.2 revealed several principles and fundamentals that should be overarching for the investigation. There can be a lot of great methods that can be used to meet modern challenges, such as increasing digitisation, but they need to be in line with the requirements to complete a high-quality investigation.

Several questions were posed to the users regarding these principles. Some of the primary principles are chain of custody and evidence integrity. To fulfil these requirements the right competence must be in place, and methods such as documentation, verification and validation are crucial to the process.

4.3.1 Verification and validation

One practical way to implement verification and validation of data is to use multiple tools to control the integrity of the data. When asked how much users trust that Hansken gives valid data, most people do trust it as per Table 2, but they use a secondary tool to verify data. 13% of the users use only Hansken when working with digital evidence, the rest use multiple tools. The top reason for using other tools are to verify data/results, the next is because of lack of support for a specific task in Hansken, as seen in Table 3.

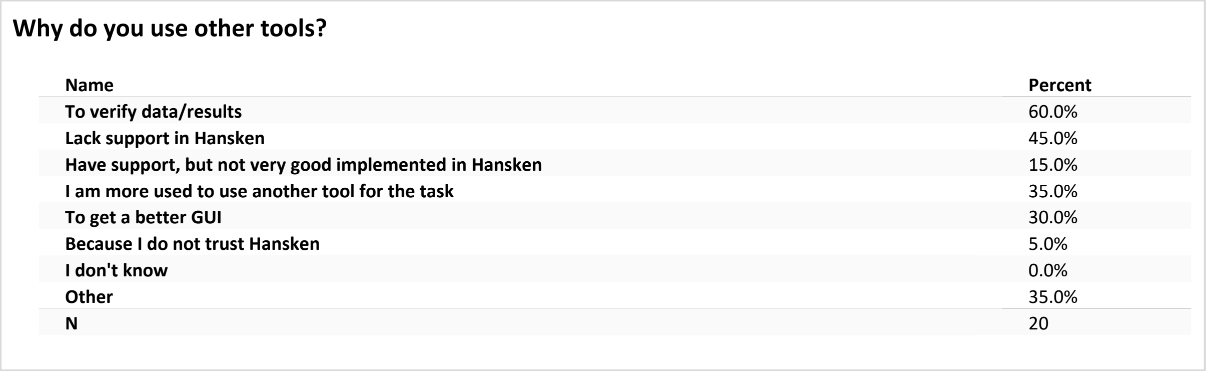
Table 2: Survey question regarding trust



On a scale from 1 to 5 how much do you trust Hansken to give you valid data?

Name	Percent
1 (I trust it enough to only use this tool and confidently present my findings in the court)	8.7%
2 (I trust it, but I regularly use a secondary tool to verify data)	65.2%
3 (Neither trust or distrust it, results need to be verified and validated any way)	21.7%
4 (I do not trust it before the data gets verified and validated in multiple tools)	0.0%
5 (Do not trust it at all)	0.0%
Only used Xiraf	4.3%
I don't know	0.0%
Other	0.0%
N	23

Table 3: Survey question regarding use of multiple tools



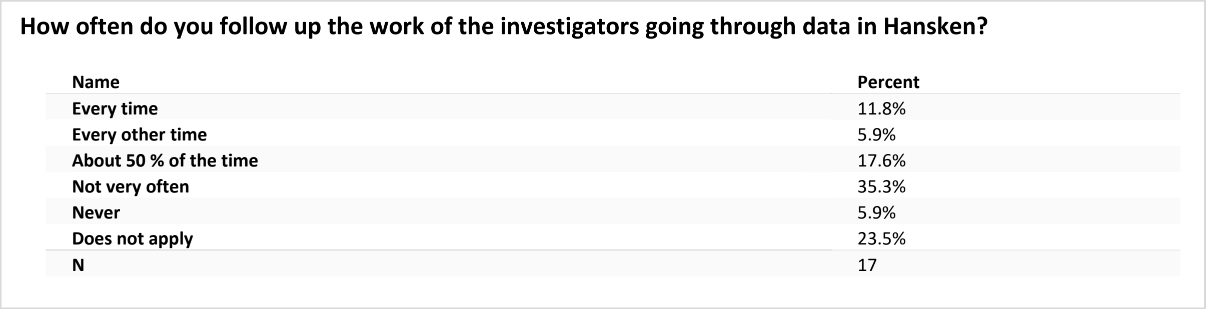
Why do you use other tools?

Name	Percent
To verify data/results	60.0%
Lack support in Hansken	45.0%
Have support, but not very good implemented in Hansken	15.0%
I am more used to use another tool for the task	35.0%
To get a better GUI	30.0%
Because I do not trust Hansken	5.0%
I don't know	0.0%
Other	35.0%
N	20

On verifying others work or having one’s work verified

Based on the characteristics of the users in Section 4.1, of the ones who had digital investigation as their occupation, 70.6% followed up on the work of colleagues, where 11.8% did this every time. 5.9% did never do this (Table 4).[11]

Table 4: Survey question regarding how regular others' work are followed up



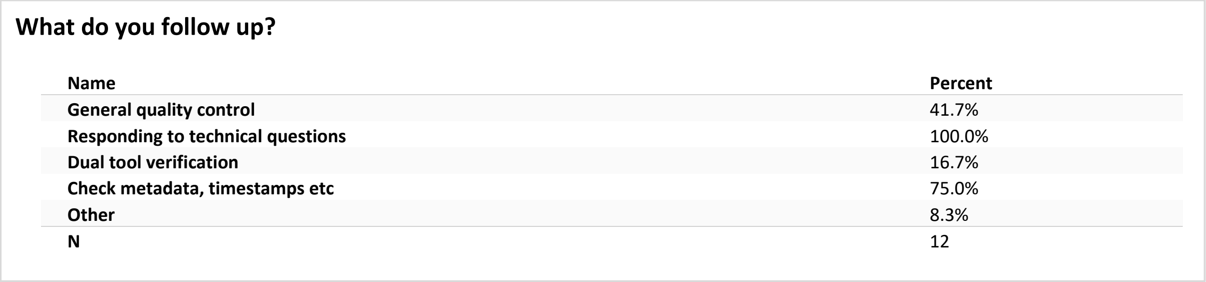
How often do you follow up the work of the investigators going through data in Hansken?

Name	Percent
Every time	11.8%
Every other time	5.9%
About 50 % of the time	17.6%
Not very often	35.3%
Never	5.9%
Does not apply	23.5%
N	17

[11] Here, users that said they just gave guidance and support (did not verify) was included as well. So, the answer “never” may also mean it is not their role, even “does not apply” should be the correct answer in that case.

The ones who did follow up was asked what they reviewed, and the top reason was with regards to answering technical questions, next came to check metadata and timestamps, and after that general quality control (Table 5). Asking the ones who did *not* follow up, 60% did not because it was not part of their profession. 20% did not have the time, and 20% felt they did not need to – for unknown reasons (Table 6).

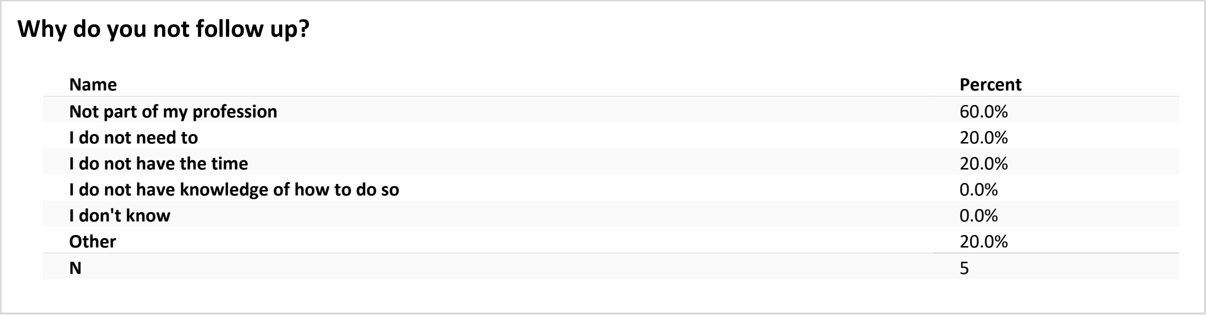
Table 5: Survey question regarding what they follow up



What do you follow up?

Name	Percent
General quality control	41.7%
Responding to technical questions	100.0%
Dual tool verification	16.7%
Check metadata, timestamps etc	75.0%
Other	8.3%
N	12

Table 6: Survey question regarding not following up

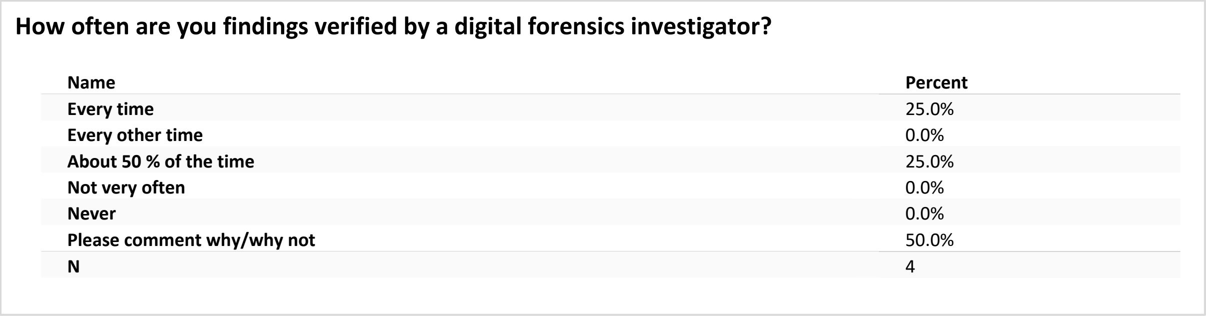


Why do you not follow up?

Name	Percent
Not part of my profession	60.0%
I do not need to	20.0%
I do not have the time	20.0%
I do not have knowledge of how to do so	0.0%
I don't know	0.0%
Other	20.0%
N	5

With regards to the ones who did not have digital investigation with additional responsibilities such as supporting others, all of them had their work verified, either every time or about half of the time (Table 7).

Table 7: Survey question regarding how regular their work is reviewed



How often are you findings verified by a digital forensics investigator?

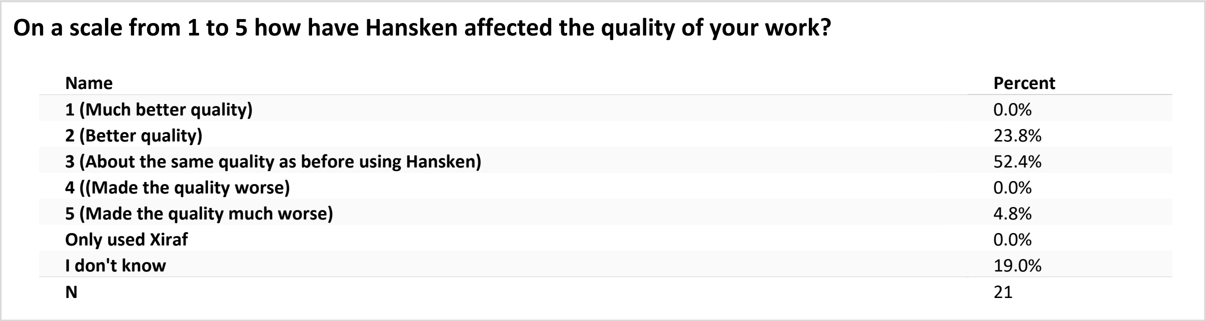
Name	Percent
Every time	25.0%
Every other time	0.0%
About 50 % of the time	25.0%
Not very often	0.0%
Never	0.0%
Please comment why/why not	50.0%
N	4

One commented that a big risk is when non-technical investigators make their own conclusions, and that results that are presented for the court should always be verified. The users were asked whether they had presented results from Hansken in court, five users had done so. When asked if the court asked something about how they came to their results/conclusions, one user had received this kind of question from the court. When it comes to the question whether the investigators feel that they are released from much of the pressure in the investigation to do other important things, such as research, development and deeper analysis; 35.3% answered yes, 58.8% said no, and 5.9% were not sure.

4.3.2 Quality and efficiency

When asked how Hansken has affected the quality of their work (Table 8), 52.4% say the quality has stayed the same as since before using Hansken. 23.8% say Hansken has made their work consist of better quality. One user said it has made it much worse, unknown why[12].

Table 8: Survey question on the users' perception of quality

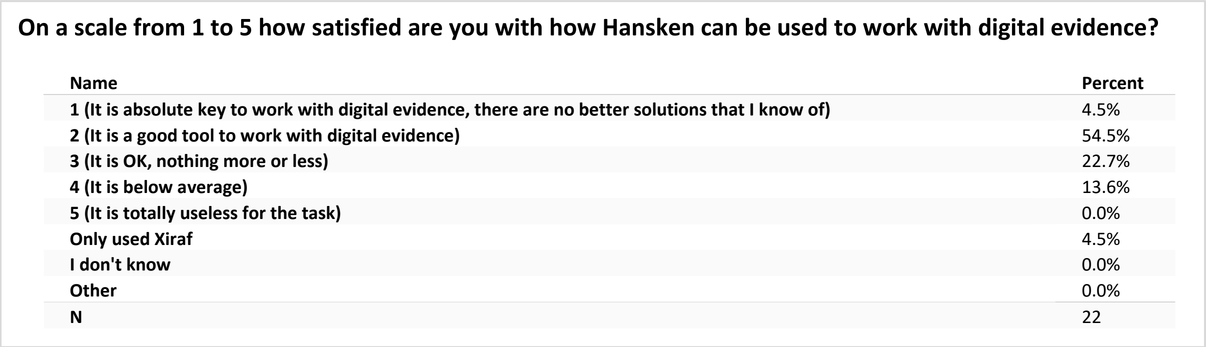


On a scale from 1 to 5 how have Hansken affected the quality of your work?

Name	Percent
1 (Much better quality)	0.0%
2 (Better quality)	23.8%
3 (About the same quality as before using Hansken)	52.4%
4 ((Made the quality worse)	0.0%
5 (Made the quality much worse)	4.8%
Only used Xiraf	0.0%
I don't know	19.0%
N	21

Most of the respondents found Hansken to be a good tool to work with digital evidence (Table 9):

Table 9: Survey question regarding level of satisfaction

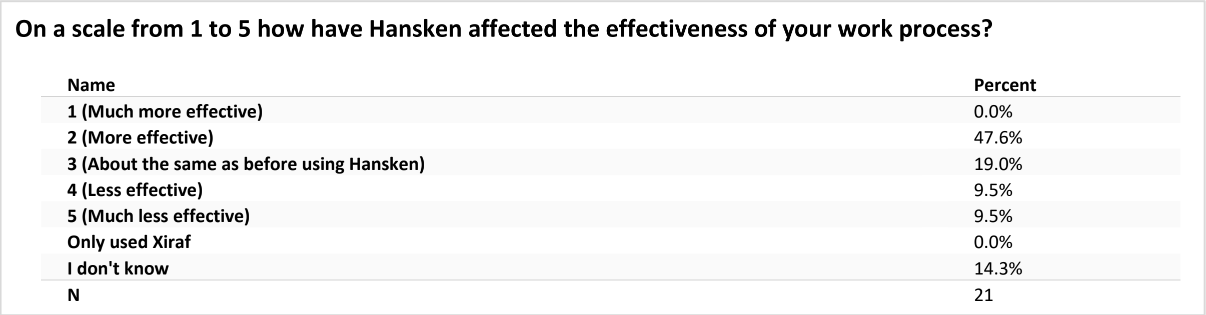


On a scale from 1 to 5 how satisfied are you with how Hansken can be used to work with digital evidence?

Name	Percent
1 (It is absolute key to work with digital evidence, there are no better solutions that I know of)	4.5%
2 (It is a good tool to work with digital evidence)	54.5%
3 (It is OK, nothing more or less)	22.7%
4 (It is below average)	13.6%
5 (It is totally useless for the task)	0.0%
Only used Xiraf	4.5%
I don't know	0.0%
Other	0.0%
N	22

Most of the respondents also experience more effectivity in their work process after using Hansken (Table 10):

Table 10: Survey question regarding level of effectivity



On a scale from 1 to 5 how have Hansken affected the effectiveness of your work process?

Name	Percent
1 (Much more effective)	0.0%
2 (More effective)	47.6%
3 (About the same as before using Hansken)	19.0%
4 (Less effective)	9.5%
5 (Much less effective)	9.5%
Only used Xiraf	0.0%
I don't know	14.3%
N	21

[12] When developing the survey, the mixed-methods design was used so that the users had the opportunity to comment underway. But this was only optional, and not all used this functionality. Because of this not all answers "makes sense" – see strengths and weaknesses in next chapter.

Some commented on performance issues and instabilities which could have nothing to do with Hansken itself, but for instance because of challenges with the infrastructure. One user commented that with regards to the implementation, the availability was prioritized from the start when the system was released, so that every police officer could get easy access from their office. But that the system has suffered from issues with bandwidth since then[13].

A lot of the users found that Hansken has a lot of potential, and that it is great to get an overview of the case. But on the other hand, some found it not suited for such tasks as incident response, because it takes time to get up and running. One concrete reason was due to the delay in the upload process.

In summary, as stated in the preamble of the chapter the main topics for the survey was opportunities and requirements in the investigation with a DFaaS system called Hansken.

The users were posed several questions based on the tasks in hybrid digital investigation process model and each was rated with an average score that shows whether they are fulfilling its potential in the DFaaS system. Almost all the tasks had a score above average. The exception was related to the *acquire data* phase where they found the process to be slow with regards to the time it takes before one have access to the data. Infrastructure related issues such as bandwidth could be one of the reasons for this[13].

Further, there were several questions posed on the fundamental principles of the digital investigation. In this regard most of the users trust Hansken to give them valid data, but they did use other tools to verify their findings.

Based on the results from the survey several topics are discussed further. The topics of implementation strategies for the DFaaS system to fulfil its potential, along with quality requirements, and what DFaaS still probably needs in terms of the right expertise is some of the focus areas of the next chapter.

[13] This was confirmed by one of the product owners, that this loss in performance did not have its root in Hansken itself, but that the system had trouble because of the implementation on the infrastructure. This is discussed further in Section 5.6.

5 Discussions

In this chapter, the research question formulated in the introduction will be worked towards answering by discussing each sub question. The goal of these sub questions was described in the introduction, Section 1.3, as to "(…)deconstruct the main problem into smaller more manageable problems, which together make the foundation of answer the main problem".

Section 5.1 starts off with a general discussion of the results from the study. Then each of the sub questions are posed one by one:

- **Section 5.2** discusses the process model in relation to the DFaaS system, and asks the question; *Are the digital forensics process models valid work models for DFaaS?*
- **Section 5.3** looks at the first requirement, which is the efficiency requirement and asks; *What socio-technical measures are suitable to attain efficiency in DFaaS?*
- **Section 5.4** looks at the second requirement, which is the quality requirement and asks; *What socio-technical measures are suitable to attain quality in DFaaS?*
- **Section 5.5** seeks to tie the measures together in a socio-technical harmony by asking the question; *How can the socio-technical system in DFaaS be balanced and what constitutes socio-technical balance?*

Last, the strengths and weaknesses of the study is in reviewed in **Section 5.6** along with a summary of the chapter.

5.1 General discussion of results

The importance of a stable foundation with a good infrastructure
It was identified from the feedback of users, experiences relating to performance issues and instabilities. This could have shaped the view of the participants, both to how they rated the methods and with regards to their satisfaction with the system. Investigating the reason for such issues was beyond the scope of the survey, but some commented it could be linked to challenges with the infrastructure. This was also confirmed by one of the product owners of Hansken, see Strengths and weaknesses in Section 5.6. This would be an important requirement if the efficiency, but also quality is to be attained within the DFaaS model

It is also very important that the methods are compatible with the application, and that the users have a healthy trust relationship and knowledge of how it works and that it is implemented correctly.

It would be a contradiction if the system slows down the investigation because of delays and bad performances with a system that is supposed to speed up the accessibility and investigation of digital evidence.

These questions are discussed further in the later parts of the discussion, where concrete strategies are recommended that ties directly in to the research question.

The importance of implementing verification procedures

As was identified in the results from the survey, with regards to the fundamental requirements of the digital investigation (Section 4.3), most users (65.2%) found the data Hansken presented as trustable in that they used a second tool to verify data. Only 8.7% trusted it to that degree of just using one tool. When the experts were asked how often they did review and control the work of others, most did not do this very often.

It then bears to ask, what about the detectives that do not possess competence on handling other forensic oriented tools? One thing is the trust the users feel in the validity of the data, another thing is actual pitfalls that were identified in Section 2.3. In addition, if most experts do not validate the work of others every time, how can the less technical oriented detectives know that the data, results or conclusions are correct, every time?

In a DFaaS system, one of the advantages is that digital investigators are released from much of the administrative overhead, and backlogs would decrease as they should have time to address others work.

Whether this is the actual case that backlogs have decreased have not been concluded in this study since there was not any luck in collecting this data, as seen in Section 2.1.2.

Scenario, part 3

Harry logged in and started to go through the data, but he had to wait for each picture to load, and the connection timed out several times. He informed CTU, and they told him they were working on it. Harry went on with other investigative activities.

Harry interviewed the subject and confronted him with the findings. The subject was very surprised and swear he had no clue that this was on his computer. He had downloaded "something", but he was convinced that is was legal material. The subject explained that he did not have very good technical abilities, and that his machine had been acting strange lately by being slow and giving random pop-ups that disappeared before he could see what the content of it was.

When asked why he had so many storage devices, he explained that he had a personal business that did video recording for real estate with drones. Harry found it very suspicious that the subject had all this "computer stuff" and not technical abilities.

CTU was informed of this, and quite pleased since they could apply a filter in the processing so that the drone recordings were skipped. The data material to review went down drastically, but because of the World Ski Championships, "everyone" was streaming the finals at their offices. Since the DFaaS system was using the same infrastructure, the lag in the system was still unbearable. Harry asked the Digital Forensics unit for assistance again, but it was denied because of capacity. CTU also had their hand full of providing assistance all over the country with the new system.

Harry conferred with the prosecutor on the case regarding the challenges, and he decided that they had enough material to give the subject a fine due to exercising negligence in being in possession of the material.

The subject did not accept the fine, and the case went to court where the defence asked for checking the computer themselves. The judge accepted this claim, since the police had not reviewed all the data. The defence found a trojan virus on the computer that had been externally communicating with an underground web site that was known for spreading illegal material, and the subject was exonerated and compensated. The news spread quickly and the police in Torskevik got massive critic.

(Continues on page 71)

5.2 Are the digital forensics process models valid work models for DFaaS?

For the DFaaS system to fulfil its potential while meeting the fundamental requirements of quality and efficiency, the process model must be compatible with this system to support these requirements. In Section 2.2.1, many of the various process models that was developed over time was reviewed. In the methodology chapter, some requirements to a process model to work with was proposed. A hybrid process model that consisted of several other models was suggested and used while developing the survey and doing the experiments in the thesis.

One way to angle the question is to see whether this model is valid for DFaaS in itself. Another way would be to use this model as a baseline for the socio-technical methods in DFaaS and see if the other parts of the socio-technical system model (machines, culture and structure) are compatible with this process model.

The first view on the question, if the model is valid for DFaaS in itself is hard to answer when only this model was used, and it was not compared with other models. The requirements to the process model, set in Section 3.1; *should be detailed, easy to read and include quality measures*, was mostly created to have a clear framework to work with that ties directly into the problem statement of the thesis. However, as mentioned in Section 3.1, the tasks in the model was taken from the Integrated digital forensics process model which was also used by the developers of Hansken. That is why it felt natural to use this as a base.

In the survey, the respondents were asked questions about the work processes which scored above average in terms to the degree of which the steps were compatible with this model. It can be asked if this score implies compatibility with the process model in DFaaS or it can be questioned how this score could be improved, independent of the first question.

To answer the first question, a better data foundation in terms of variety (multiple tools and models) and quantity (more respondents) would probably have been required. This is outside the scope of this thesis. The second question is more productive to work with and does not require the same degree of precise data. “We admit the data could be inaccurate, but we want to improve it anyway”. This is illustrated in Figure 20: How can the methods average score be improved? And how can the results be skewed towards “Absolutely”?

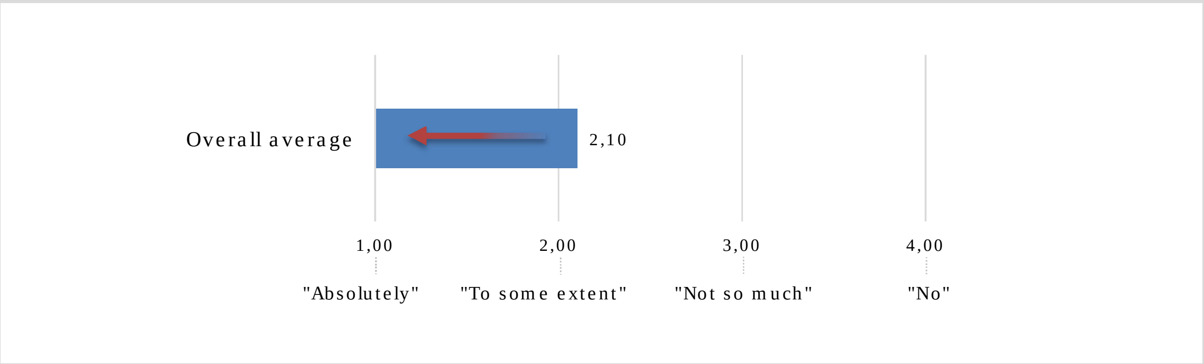


Figure 20: How can the methods average score be improved?

5.3 What socio-technical measures are suitable to attain efficiency in DFaaS?

After working with the question in the last section, the author decided to look for ways to improve the DFaaS system with a basis of the findings in survey.

In this section, the phases of the hybrid process model will be used as a baseline to try and improve these in the DFaaS model based on the results from the data identified in Chapter 4. The two last phases "Analyse data" and "Evaluate hypotheses" are clustered together, because of their tight relationship. The socio-technical approach will be used to structure this by using the four parts; methods, machines, culture and structure. The goal is to give recommendations for each of these in each phase of the process model.

To give a short recap of the process model, it consists of:

1. Acquire data
2. Explore and examine data
3. Analyse data
4. Evaluate hypotheses

Since the two last cases are merged, this is the phases used:

1. Acquire data
2. Explore and examine data
3. Analyse data and evaluate hypotheses

Several points will be repeated throughout the phases, this is because they would apply at different places in the process.

Disclaimer and limitations
The points given with regards to improvement should be general for all DFaaS systems, but since this study focused on Hansken in the Netherlands it is natural that the ideas came with regards to feedback of this system. This list is by *no means complete* and should be a dynamic process. Its intention is to give examples on how the different part of the socio-technical system are connected.

When developing a DFaaS system one should take into consideration how Hansken is functionally built with its advantages that are integrated already. This is not discussed here, but needs to be considered. Also, it is important to state that multiple of the proposals mentioned *may* already be implemented into Hansken. These were not conferred with the NFI (developers of Hansken), and they are based on the literature review, especially the state of the art in Section 2.4.

Suggested measures

1. **Acquire data**
   This is the phase that has the greatest potential since it had the most negative score of all the phases. Users rate the accessibility to data early in the case below average. One of the reasons is because the bottleneck in the initial phase of the investigation which is the time-consuming imaging process and uploading of data, which could have its root in infrastructure challenges and therefore affected the users' opinions. The potential for latency and stability issues was also identified in (6).

a) Methods[14]

Several methods could be explored to augment the acquisition phase (see Section 2.4 where they were reviewed), some of which are:

- Triage: By conducting a preliminary survey that involves the rapid review of many potential sources of digital evidence, one can quickly identify those items that contain relevant evidence. As a consequence, the number of devices to image could be greatly reduced.
- Remote data acquisition: Enables remote acquisition and transfer of the image from any seized device directly to a forensic laboratory for analysis.
- Data reduction: With regards to acquisition, selective imaging can be used to reduce the amount of data that is included in the imaging process. This includes using filters/conditions to concentrate on potentially relevant data.

Methods such as triage and data reduction must be balanced dependent on the type of case. One of the goals for the digital investigation is to minimize case lead time while maximizing the coverage of traces – two quite contradictory desires. In some instances, all the data in a case needs to be reviewed and evaluated due to the risk of missing crucial data/information. At the same time, there is a risk for losing evidence if everything is included into the case, either by "drowning in data", or loosing volatile data because of the additional time required to access the data and commence the investigation. It is however, one of the strengths of the DFaaS system to handle large amounts of data.

b) Machines

Considerations for the machines are the following:

- Integrated acquisition: The tasks related to the acquisition could be part of the machine/tool. Today in the Hansken system, the imaging process is a separate prerequisite step. If this is implemented into the process and become more automated, the case lead time could be minimized.
- Expandable and future proofing: If implementing further methods, there will need to be created support for this in the system.

c) Culture

For the culture it is important to:

- Increase efficiency of capacity: By having a better integration, as was exemplified in the last point, the human capacities are distributed better within the system. For instance; instead of digital investigators doing the imaging of data, the system administrators[15] can do this as a natural step in the DFaaS process. This will specialise individuals more, allowing them to concentrate on the tasks that require their enhanced competence.

[14] It is important that the personnel conducting the methods have the competence and comprehension. This is further discussed in Section 5.5.

[15] This obviously requires these personnel to have the appropriate training. The point is to specialise people more, so that the specialists are not becoming jack of all trades. This was also pointed on as a strength with the DFaaS system by the Hansken developers (6).

d) Structure

For the structure, the recommendation is:

- Stability, accessibility and speed: To deal with the potential for latency, the infrastructure must be adept at meeting such demands. An effective investigation requires favourable I/O speeds and good bandwidth. It would also demand a stable access, with a predictable and favourable up-time.

2. Explore and examine data

The questions in the survey with regards to this phase, focused on the level of data completeness and data purity – the degree to which relevant traces are included and whether data noise, such as system files, are filtered away. The scores could be improved, especially with regards to data purity. The feedback was that it could be hard, especially for a non-technical individual, to distinguish the characteristics of files such as the level of completeness, the source of files/devices and the degree of relevance/non-relevance.

a) Methods

The methods that could be added are:

- Data mining (computational forensics): Comprises of using a combination of knowledge discovery techniques to find relevant and useful data. Examples are statistical modelling, machine learning, pattern recognition, data visualization and database processes. The investigators would have access to the relevant data quicker and would save time, limiting their review of non-relevant files.

b) Machines

With regards to the machines, they should have/be:

- Stability and speed: To have a stable and rapid processing of data it is important to mitigate bugs and have the adequate hardware. There should be continuous controls for bugs in the system and the hardware development should be dynamic to meet increasing demands.
- Expandable and future proofing: If additional methods are to be implemented in the future, a requirement is that there is support for this in the system.

c) Culture

The culture should promote:

- Competence and collaboration: To discern between the different type of traces it is important that the investigators have at least some basic technical knowledge[16]. It should also be easy to ask more qualified investigators when in doubt, and the work should always be verified by others (peer review). By having too little knowledge the risk for misinterpreting traces and focusing on the wrong data will always be present, as discussed in Section 2.3.

[16] What this specifically implies is outside the scope to discuss.

d) **Structure**

The structure should have:

- Stability, accessibility and speed: The infrastructure should promote up-time and mitigate latency to support this phase.

3. **Analyse data and evaluate hypotheses**

These two phases were merged since they are tightly woven together (the analysis is carried out with a basis in a hypothesis and findings may confirm or deny the hypothetical question). Some users in the survey had some difficulties with the interface and there is room for interpretation errors, as there are in all digital forensics tools as elaborated on in Section 2.3. The survey also identified some circumstances that is often seen in "all-in-one" tools, like missing features that are found in specialised tools.

a) **Methods**

Methods that are proposed for this phase are:

- Intelligent analysis methods (computational forensics): As identified in Section 2.3.3, growing research is being conducted with regards to automate analysis more. It is worth exploring new methods, but as was identified earlier, it still requires competency and the right expertise to interpret the meaning of data, to evaluate and to make conclusion. Intelligent methods to this phase should be treated as supplementary aid rather than a replacement of manual expertise.

b) **Machines**

For the machines, the considerations are:

- Apt user interface: Having an interface that promotes the analysis process is vital to the investigation. It should be clear and without doubt on what abstraction level the user should see depending on the competence. It is important to this regard to know that there is a discrepancy between giving less experienced investigators more high-level interpreted traces, and the fact that this can increase risk of errors in the data (see Section 2.3.1).
- Expandable and future proofing: To have the opportunity to implement additional methods, there will need to be support for this in the system.

c) **Culture**

The culture should have:

- Competence and collaboration: Users must be able to understand the meaning of what they see, and experts will have to verify the data and results that is interpreted by less experienced users. This is addressed as an important matter in Section 2.3.

d) **Structure**

For the structure, the following is proposed:

- Stability, accessibility and speed: The infrastructure should promote up-time and mitigate latency to support this phase.

5.4 What socio-technical measures are suitable to attain quality in DFaaS?

Improving the efficiency without the quality concern would be quite meaningless if the investigation requirements are to be fulfilled, although it was touched upon, such as the competence in the culture. This would count for both efficiency and quality.

In this section the focus is purely on the quality of the investigation, without any efficiency requirements. How does these relate together then, one might ask. That is the topic in the subsequent Section 5.5.

The same socio-technical analysis is applied, where each of the socio-technical parts of the DFaaS system are enforced with fitting measures.

These measures are mostly based on findings from the literature review and partly from the results of the survey. These should be seen as general to DFaaS, and the specific system that was focus of the survey Chapter 4; Hansken, may or may not have these measures implemented. This was outside the scope of the research to investigate further.

This list is, like in the previous section, *by no means complete*, and should be treated as a suggestion rather than an absolute. The intention of the author is to provide examples in both categories (efficiency and quality) and then compare their compatibility with other by giving further measures.

Suggested measures

a) **Methods**

Here methods are listed that are specific for the quality control throughout the investigation:

- Documentation: Should record all information relevant to the investigation process, i.e., describing the state of systems when data is collected. Including what tools and methods was used, and it should include all findings – both relevant and findings later deemed irrelevant.
- Evidence preservation: Should maximize evidence availability and quality, while maintaining the integrity of the evidence during the digital investigation process. Concrete examples of methods are to use write-blockers[17] in the acquisition phase and always incorporate read-only mode against source data. If this is not possible due to running systems, cloud etc., the points with regards to documentation would be especially important.
- Verification: Includes using multiple tools and sources, checking data manual at different levels, hash comparison, testing of hypotheses/tools and peer review.
- Planning and dynamic approach: It is important to have a plan for the investigation that includes tasks, hypotheses and deadlines[18]. It is also important to adjust the measures accordingly to the case type.

[17] Only possible in post-mortem acquisition

[18] The Attorney General emphasize the importance of planning to have a greater quality of investigations, and making deadlines as a concrete measure to mitigate potential backlog of cases (27).

b) Machines

To mitigate for errors in the machines the following is proposed:

- Integration of quality measures: Make verification easier of sources and between tools, plus have documentation as a feature and detailed logging to support the chain of custody.
- Fulfil Daubert standards: Independently testing of reliability, potential error rates must be available, and define standards for tool handling and have acceptance in community.

c) Culture

To better equip the culture, the considerations are:

- Competence and training: Have the right competence and expertise at the right place in the system/process. Have clear defined requirements to the specific tasks and set minimum requirement to training.
- Quality oriented/ethics: Define quality oriented goals and have focus on ethics and overarching principles such as objectivity, independence and impartiality. Work to promote trust in the relationship with the public and investigated parties. Focus on potential pitfalls in the investigation such as confirmation bias, which should be part of the theory in the training.
- Increased cooperation: Define openness and communication as clear values. Have clearly defined roles to the investigation.
- Leadership and management that is fully integrated in the investigation and quality oriented, by having the knowledge and experience of potential pitfalls, such as the potential for errors in data and in the interpretation of data. Further has knowledge of how this can play out later in later stages of the case, such as trial, hearings etc. The management also need to make the investigation goal oriented, with a discernible supervision[19]

d) Structure

In the structure, the following are proposed:

- Quality oriented routines: Implement routines that seeks to promote the execution of investigative tasks, by controlling the quality of the work and make it integrated in the organization. This is directly connected to the leadership and management.
- Education: Needs to be integrated into the structure, with compulsory training and continuous education.
- Cohesive: The structure should promote collaboration and create close proximity between developers, administrators and users.
- Laboratory accreditation: To support a high standard of instruments and tools there should be some basic requirements to the equipment used in the investigation.

[19] The Attorney General stresses the importance of this to uphold the quality of the investigation (27).

5.5 How can the socio-technical system in DFaaS be balanced and what constitutes socio-technical balance?

This section will seek to tie the different requirements of efficiency and quality together, and when answered will provide a better foundation to answer the research question.

What does it mean to have socio-technical balance in a DFaaS model?
A continuous topic throughout this thesis has been regarding the importance of good relationships and harmony in the socio-technical system. The goal has been to strengthen each part of the model; the methods, machines, culture and structure. This is not an individual process, it is much more comprehensive and conditional. An example would be if we were to introduce a new machine, which could be a physical computer, an application or a distributed server node. Our methods will need to be *compatible* with this instrument, the culture will need to *adapt* to use it, and the structure must be built to *serve* it in a beneficial fashion.

In the experiment, the focus was on an established DFaaS system in the Netherlands called Hansken with a concrete group of users; Dutch law enforcement employees in a Dutch law enforcement organization. A process model based on established methods for digital forensics, was used as a baseline for the work process in this system. This could as well be any DFaaS model in any other country and could have provided completely different results, including a different culture in a different structure, potentially using the same tool, but with different hardware.

The ultimate goal for a DFaaS system could be argued is the same; to strive towards assisting the investigation in an effective, but safe way. These two requirements; efficiency and quality can be conflicting and hard to reach both. That is why balance between these are required.

The motivation and goal for a DFaaS model
Some of the benefits of using DFaaS were identified in Section 2.4.1 and 2.4.2. The Hansken developers summed it up fittingly (31, p. 2): "Our main goal is to provide a service that processes high volumes of digital material in a forensic context and gives easy and secure access to the processed results". They identified three forensic drivers which were "minimization of case lead time, maximization of the trace coverage and specialization of people involved".

The goal is in other words to increase the *efficiency* of the investigation and increase the *accessibility* of the data by giving investigators early access to it. Still facilitating *collaboration* between case investigators and digital forensics investigators and utilising *expertise* better. At the same time the *principles of the digital investigation must be fulfilled* to make sure it is in symbiosis with the rule of law. This is also where the paradox of automation comes into play; we need systems such as this to be able to keep up with the technological evolution, but implementing this requires the system to be secure and trustworthy.

How to attain balance and harmony in the socio-technical system?
The measures that were identified seek to improve the scores of the users with regards to the digital investigation process and the quality of the investigation. This needs to be balanced against the rest of the socio-technical model, or else the implementation would

be contradictory and quite pointless in a socio-technical system since the system requires balance to work properly.

If we are to implement, say data mining, the method itself must be ***usable*** and ***versatile*** for the users, and the users need to have the right ***competence*** and ***comprehension*** of how the functionality and results of this method plays out in the investigation. The method must be ***tested*** and well ***documented*** to have the right ***support*** in the structure of the organization, and it ought to be feasible to ***implement*** it in the machines which must be ***compatible*** with the method. If taken a step further, implementing the method in the machines requires that this is ***accomplishable*** to do in the structure which needs to have the ***capacity*** to do so. For the culture to successfully use the method it would need ***capability*** and ***skills*** to operate it in the machines; which should be ***manageable***, convenient and easy to use. The culture would also need to be ***organized*** in a way that promotes ***cooperation***, and this would require that the structure is ***supportive*** by being a ***stable*** and ***reliable*** foundation for the investigation team.

In the following, the relationship of the different parts of the socio-technical system are listed with bullet points on how the harmony between these can be accomplished. The important aspect is that they are all related.

Methods	⟷	**Culture**
Must be usable and versatile		Must have comprehension and competence
Culture	⟷	**Structure**
Must be organized and cooperative		Must be reliable and have stability
Structure	⟷	**Machines**
Must have capacity		Must be accomplishable
Machines	⟷	**Methods**
Must be compatible		Must be implementable
Methods	⟷	**Structure**
Must be tested and documented		Must provide support
Culture	⟷	**Machines**
Must have the capability and skills		Must be manageable

In conclusion, to promote balance and harmony in the socio-technical system the *methods* need to be usable, versatile, implementable, tested and documented; the *culture* needs to be comprehensive, competent, organized, cooperative, capable and skilled; the *structure* needs to be reliable, stable, have capacity and support; and the *machines* must be accomplishable, compatible and manageable.

5.6 Strengths and weaknesses of the study

Generalisation from one DFaaS system
The thesis with its problem statement is focused on DFaaS in general and individual variations in selected DFaaS systems may not be representative for DFaaS in general. Hansken was chosen as a concrete example of a DFaaS system that has been in use over multiple years, deployed to the law enforcement in the Netherlands. Conclusions from the study will be based largely on inductive reasoning from this system to DFaaS in general.

Study could be generalised to similar studies
This study could probably be generalised to other investigation models/-systems as well. A socio-technical approach is a sensible measure to use when analysing the compatibility or effect of new technologies, or general work methods, measured against the users in the system.

Generalisation from users
The selection of users for the survey was done by deploying via user databases that had users from all over the country on different levels in law enforcement (e.g., local, regional or national). The selection was chosen based on nonprobability sampling, so there was no way to control the variety and randomness in the user mass.

This could be a weakness in that the generalisation could be from just a subset that is not representative of all the Hansken users in the country.

Anyway, the sample show that there was a fair variety in terms of the characteristics of users, which is discussed further in "External validity", later in this section.

Limited size of the sample
The number of actual respondents in the survey was 28 people. This was around 10% of the approximate number of users that received the invitation to participate in the survey, which totals about 1% of the total amount of Hansken users. This was right on the threshold that was set in Section 3.3.1.

The small number of actual respondents could introduce less randomness than if the number was higher. This means that each response will have more ramification on the output of the combined answers. To mitigate the potential for the responses to be skewed towards the extremity regarding opinions, it is important to look for outliers among the respondents and decide whether this could affect the average results more. The use of a mixed-methods study with an embedded design should compensate for this, since possible explanations, in the form of comments, would enrich the responses. However, since comments was optional, this was not always possible, as displayed in Section 4.3.2, footnote 12.

Possible explanations of a limited sample size could be:

- Distance between target group and researcher: The target group was in another country than the researcher and held no familiarity to the author. They only had limited written information about the background of the researcher and the motivation for the study. In addition, most of the publication of the participation in the study was done via the product owners of Hansken, not the researcher directly.
- Anonymity of participation: There was no link between the responses and the respondents. This meant there was no way to reach out with follow-up questions or reminders in any way.
- Time required to participate could have been underestimated: It was estimated that it would take approximately 10 minutes to complete the survey, but this number could be underestimated depending on how quick the respondents gave an answer.
- Possible unclear what the profit/motivation was: The purpose of the study was described as (short version): "To investigate how the digital investigator can maintain the quality on the investigation with this model (DFaaS)"[20]. The researcher personally felt that this is a very important question to ask, but there is no way to know if the respondents feel the same.

One way the numbers probably could be increased was to send a reminder to the non-respondents. But this was not feasible, due to the anonymity of the participants.

Opinions and biases

The study is not meant to be a review of Hansken, the author has no previous experience with such a system and has not used Hansken. The data is based purely on other people's opinions balanced against the authors own interpretations. However, by having this distance, could strengthen the neutrality towards the system and introduce less of a personal bias.

The author has developed a critical view towards new methods over the years, working as a digital investigator. Not as a cynic, but more of a healthy questioner. Having seen erroneous data in output from tools and found that the expert competence has been crucial to multiple cases when case detectives have browsed through digital evidence. A sceptical view can be positive when balanced against the real need for new methods and tools to face the challenges posed by the new developments in technologies, which also give us new opportunities.

When it comes to bias from the respondents, the user mass should probably have been larger, and the selection should be more random than just users spread around the country. One part of the country could have been overrepresented in the sample, and bias from that group would have the potential to affect the total mass.

Further, most of the literature on Hansken is from the developers themselves. This could have the potential to skew the information towards the positive. However, their written articles (6, 31) along with conversing with the developers they appear to be rather humble,

[20] This was the initial research problem, and since then it was expanded quite a lot.

the development is very transparent, and they are continuously requesting feedback and suggestions from colleagues.

On verification

In the survey, only the users that did not give guidance, support or that verified data, got the question if their work was verified and the frequency of this. It would be especially important that the less technical users have their work reviewed, but more experienced users should not be excluded from this peer revision. Based on this, the more technical users should have probably been queried regarding this matter, to get a broader picture overall.

External validity

In Section 3.3.2, the requirements for the study to have both internal and external validity was identified. The internal validity in this study was also further explained (multiple sources, triangulation etc.) in that section.

Here the external validity will be identified with regards to the characteristics of the users in the sample (see Section 4.1 for the actual numbers).

When it comes to the positive side for external validity, the study should fulfil this requirement since it was using a real-life setting, based on first-hand experience from the users themselves. The mix between national and regional police association was quite even and there was a good mix between users with little experience with Hansken and more experience. Most had worked with digital evidence before using Hansken, so it is more feasible to measure the effect of Hansken.

On the more negative side with regards to external validity, the sample had the male gender dominating, however gender should not have so much to say in this study. In addition, most of the users had more "digital" competence than those with "tactical" competence, and it consisted mainly of more experienced law enforcement personnel. The degree of expertise and experience could probably be more even between experienced and non-experienced, if that is more representative of the actual users in the DFaaS system.

Identified measures are not necessarily exhaustive

It is important to mention that the measures identified, that seeks to improve the efficiency and the quality of the DFaaS system, is by no means complete. This process is very dynamic and expandable. This is meant to give the reader an example highlighting the importance of systems thinking[21], especially when implementing something like DFaaS.

Technical details and security/privacy issues not in scope

Some of the topics that was left outside of the scope could be important to answer the research question fully. This had to be done to make the amount of work manageable. This could also be related to the fact that the research question is very broad. As stated in the scope the focus was on the people involved in the investigation (digital investigators and detectives) with a DFaaS model. The question if more technical details and

[21] A school of thought that focuses on recognising the interconnections between the parts of a system and synthesising them into a unified view of the whole (117).

security/privacy issues should be included, and then left out if it proved to not be relevant, cannot be answered. This will be proposed for further research.

Temporary issues in infrastructure could affect answers

One of the product owners of Hansken that helped with sending out the invitations to the survey confirmed that there were issues with implementing Hansken on their infrastructure. He explained that because of this, some users might have found Hansken to have a slow performance. This was not due to the performance of Hansken itself in most cases, but due to the infrastructure.

In summary, several measures to how the socio-technical system in a DFaaS investigation model could be improved was identified in the above discussion. To sum it up, an illustration is presented that depicts what promotes socio-technical balance in DFaaS – Figure 21 (see Appendix E: Socio-technical balance in DFaaS, for a greater resolution).

The illustration shows the socio-technical model presented in the methodology (Section 3.1) with the addition of the measures that could increase the efficiency (blue text) and the quality (red text) of the socio-technical parts of the DFaaS model, and also what constitutes balance (green text) between the subordinate technical and social parts.

Again, as stated under the strengths and weaknesses, this is just meant to give an example and it is not a blueprint. The process needs to be dynamic. What this tries to establish is the importance of the different parts in a system to communicate better.

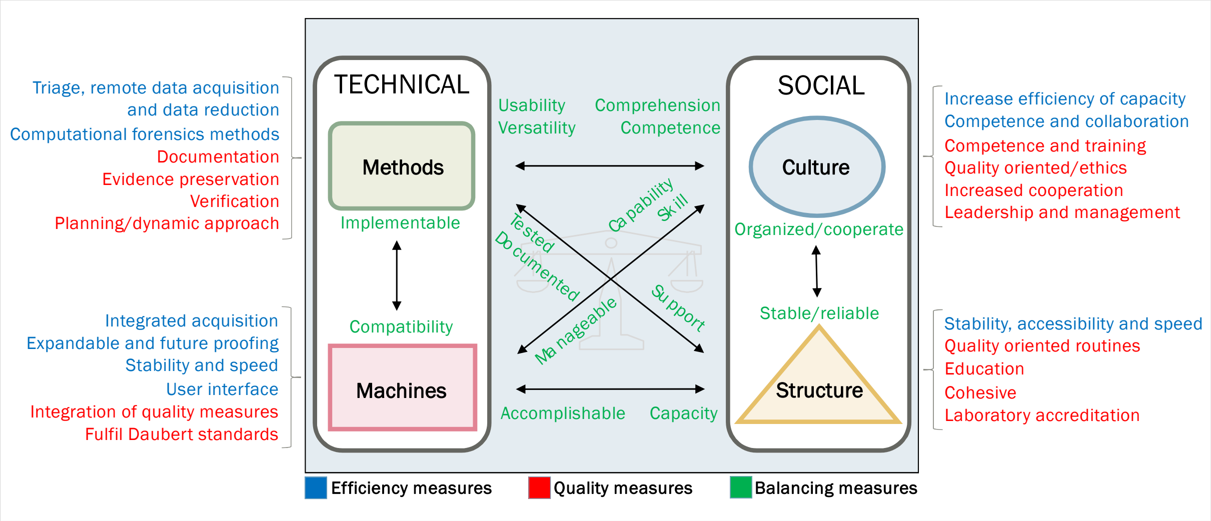


Figure 21: Socio-technical balance in DFaaS

6 Conclusion and further research

The last chapter continues the discussion from the previous chapter with focus on the research question. Using data from the experiments and results, together with the discussion of the different sub questions, the goal with this chapter is to provide a conclusion and discover what this implies in theory and practise.

This chapter is organized the following way: First the research question is answered along with a summary from the study. The theoretical implications are presented in **Section 6.1**, then heading over to the more practical considerations in **Section 6.2**. In this section, an example of a framework for law enforcement agencies is roughly sketched, which is one of the topics that could be built on further along with other suggestions to study further is given in **Section 6.3**.

Conclusion

This thesis set out to investigate the following research question:

How can law enforcement agencies use Digital Forensics as a Service to meet the modern digitisation challenges with focus on fundamental investigation requirements?

Based on the findings from this study, by combining quantitative and quantitative data in a mixed methods study, consisting of survey research, literature review, interview and having access to a test report, the following can be concluded:

The main takeaway is *balance*. DFaaS improves efficiency in the investigation with increasing automation and computing power. In addition, more actors in law enforcement agencies have access, such as detectives that do not necessarily have the technical expertise. This must be balanced with necessary quality control of the investigation. The paradox of automation is even more applicable to DFaaS than in the traditional investigation model and demands to be considered. The consequences of this can be mitigated by implementing the expertise in the right places of the process chain, together with close cooperation between digital experts and non-experts. A practical example was displayed on how DFaaS may be used to meet digitisation challenges and how errors can occur if not used accordingly. Section 6.2 will conclude the scenario by giving an example of how errors can be mitigated, and quality uphold.

The infrastructure is also crucial for DFaaS to fulfil its potential both in terms of implementation, and for the system to scale and be expandable; else the implementation would be quite contradictory.

Whether there exist a fully compatible process model for DFaaS, or if this has to be developed, should be studied further, as well as some additional proposals shown in Section 6.3.

In summary to the conclusion, *digitisation* creates both opportunities; such as new technologies and more trace sources, and challenges; such as handling capacity and backlogs for the *law enforcement*. *Digital Forensics as a Service* comes in wake of the digitisation and is an opportunity for *law enforcement* to meet the actual challenges caused by the digitisation. The *investigation requirements* are an absolute for the *law enforcement*, that governs the work so that requirements such as quality and efficiency is met.

Because of the *investigation requirements,* several new challenges are introduced, both with the digitisation and the Digital Forensics as a Service system. It demands higher accuracy in the investigation which is fulfilled with measures such as data validation and verification to uphold the evidence integrity and chain of custody requirement. It requires additional competency to meet the new trends and for utilizing new technologies. Law enforcement agencies needs to take advantage of what the modern technology brings to the table, which releases them from much of the manual and arduous tasks. But at the same time, the competence requirements and the degree of control of the processes must be evenly matched.

This thesis set out to evaluate DFaaS as a possible solution for the modern digitisation developments which are causing increasing backlogs for the law enforcement agencies. In this evaluation it was proposed several measures to how the quality and the efficiency of such a system could be increased in a digital investigation.

When new technologies and methods are introduced in a system, it is very important that this is balanced with the different socio-technical parts of the system. This is crucial to the quality of the investigation and ultimately to the rule of law.

Digital Forensics as a Service should not be treated as a solution to the quality control of the investigation in itself. Several measures need to be considered to have a successful implementation. This includes using multiple verification procedures, having tested the methods and tools according to industry standards, having the right competence, capabilities and skills to manage the tools and using the methods correctly. It needs to be managed by an aware and competent leadership and having a supportive and stable structure with enough capacity.

The proposed measures have the intention of creating harmony in the socio-technical system, both in terms of enhancing each parts of the model, but at the same time strengthening its implicit relations. The importance of this being a dynamic process should also the emphasized, since a change in the system needs to be considered holistic for the change to be in accord with its intention.

The paradox of automation is when the work processes are increasingly automated which means less human intervention, but the human control is also increasingly important. A model such as DFaaS will release some of the manual labour in the investigation, and maybe even more in the future. Notably this is good when the different skills in an investigation gets deployed to its belonging fields of knowledge where the competence can grow in the right direction.

Pitfalls in a digital investigation such as incompleteness, inaccuracy, misinterpretation and investigator bias, are still present even if implementing something like DFaaS. The risk for errors increases when more of the raw data are abstracted to a higher level via multiple levels of trust. The potential for tool related errors increases, and so does the potential for misinterpretation when more of less experienced investigators are involved in the digital investigation. But by utilizing the advantages such a model creates for the investigation, many of the potential errors could be mitigated to a level that would be tolerated in support of the rule of law.

6.1 Theoretical Implications

In Chapter 2, the theoretical foundations for the research were reviewed. The findings in this study supports the literature on the subject, but it also provides for additional research data that supports several important points. The study of users in a DFaaS system, provided concrete examples of practical experiences with the system. To the authors' knowledge this kind of study has not been conducted previously.

In this section several points that should be emphasized as theoretical implications for the field are listed.

Digital Forensics as a Service implementation and operation must adopt a socio-technical approach

To use the right approach when implementing DFaaS is crucial. One of the issues identified in the study was that too much focus on availability, and too little focus on performance would negate one of the drivers of the system, which is efficiency. Too much focus on performance, could on the other hand decrease stability. By just focusing on the efficiency, it is easy to forget the quality requirements, such as validating data and results. All these conflicting views increases the difficulty of implementing correctly.

However, by using a comprehensive approach where all the different parts of the system (methods, machines, culture and structure) are carefully being focused on, the risk for "over focusing" or "under focusing" will be mitigated.

The same accounts for the operation where the link between humans and machines are the priority, not just the machines or the humans. Humans need the right competence, and by letting the machines take over parts of the work by using more automation, it is crucial that the human control is overarching, and the right expertise is inserted into the right place in the investigation process.

The points made in Section 5.5 emphasizes the importance of having this synergy in the system.

Digital Forensics as a Service requires a compatible process model

There is a call for using a process model that is easily applicable, with a high degree of compatibility with DFaaS. The thesis set out to investigate this and ended up using a model that merged different models. Measuring the compatibility of different models to DFaaS was outside the scope for this study and should be researched further.

The use of an appropriate process model is crucial to the quality and efficiency of the investigation. It is also important that the processes being carried out in the DFaaS system is connected to the other phases of the investigation, such as planning, incident response, reporting etc. The process model should clearly define where in the process the DFaaS system starts and ends, and it should be clear who has got the different roles and functions. That means it would be easy for the investigation lead/management to differentiate on who has got the responsibility of each step. For example, that the detective can complete the content analysis (identifying and documenting relevant content information) while the digital investigator with the right expertise verifies the data and results, and does the digital evidence evaluation (determining accuracy, causation, linkages, spoliation and meaning).

There could be some dynamics in terms of case type, but there must be some minimum requirements of the investigation process. Quality control cannot be optional.

The paradox of automation has implications for DFaaS

The paradox of automation, which is the title of this thesis was explained in the introduction and the theoretical foundations, but it bears to repeat; *the more efficient the automated system, the more crucial the human contribution of the operators. Humans are less involved, but their involvement becomes more critical.* This statement directly supports the socio-technical approach. By strengthening the technical system, the social side must be equally enhanced.

Digital Forensics as a Service automates a multitude of the tasks in the investigation process. In the most extreme case this implies the detectives logging onto their computer, enters the case that is processed in the DFaaS system, starts searching for indicators that supports or refutes the hypothesis, makes a report based on the results with their own conclusions, sends it to the prosecutor, and finally goes to court with no questions asked how they came to these conclusions.

There is no reason for the detective not to trust the data that is presented in the system, after all, competent IT administrator have made it available and confirmed the process to be acceptable. Maybe the IT administrator just barely pressed a button in the application, while somebody else had already acquired the electronical devices that was seized earlier by another person. The detective has no qualification or skill to spot possible errors in the data.

The example above is an extreme one, but none the less realistic. A system that releases some of the administrative overhead of tasks is in essence very positive. By applying scripts that are developed in house that automates tasks that would take forever to do manually, and that would tire out the human and possibly introduce higher risks for error. But a system that releases the responsibility of humans, should create more scepticism. If DFaaS does this or not, is probably not the right question to ask, the more important aspect is how these systems are implemented and operated. The competence requirement is probably more important than ever and that the competence is integrated in the whole investigation chain with special expertise applied at the right place, and at the right time.

Is the term "Digital Forensics as a Service" appropriate?

Based on the findings in this study and through the previous discussions and conclusions, there should be room to discuss why this term should be considered changed. The "as a Service" models were reviewed in Section 2.4.1 and it describes different service models that offer something to the user that is hosted in the cloud, like an application.

The term Forensics as a Service emerged somewhere around 2011 (102) and variations of this have been used frequently since then. It is hard to pinpoint exactly what it implies, as there are different variations described in literature (Section 2.4.1), but the denominator should be that many of the tasks involved in digital forensics procedures such as processing of data in being distributed via a cloud like infrastructure. Processing is in essence to take forensically acquired disk images as input, and running a variety of tools and techniques that extracts and carves for data and information in different formats, which is being facilitated for the investigator.

The issue is that digital forensics includes so much more. According to the definition identified in Section 2.2 it includes the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence. Or according to the hybrid model used in this thesis; the act of acquire-, explore-, examine-, analyse data, and evaluate hypotheses. In Section 2.3.3 previous research showed that it is not feasible to automate analysis yet because it requires some degree of critical thinking and implementation of the scientific method. The examination phase is more susceptible to automation than analysis.

The study in this thesis of the DFaaS system in the Netherlands showed a high degree of flexibility, and different kinds of access levels. It could supply access via an Application Programming Interface (API) that the digital investigator could run scripts against to customize the process, but a more common use would be to access the final processed data and perform searches and visualise data.

It was further shown in this research that to create a socio-technical balance in the DFaaS system, a multitude of measures must be implemented, and a great deal of these are manual procedures, such as documentation, evidence preservation and verification. In addition, the right competence and skills must be present to fulfil the investigation requirements.

Based on all this, it could be that the term Digital Forensics as a Service could be misleading in that only parts of the digital forensics discipline are being offered as a service. By using the term, it can lead to an excuse of not implementing the necessary procedures to meet the investigation requirements.

It could also be argued that there should be different terms based on the use cases, examples of suggestions for this are:

- "(Digital Forensic) Traces as a Service", which describes what is being served to the user. Specifies end-user use case, such as a detective accessing data that has been facilitated via processing.
- "(Digital Forensic) Processing as a Service", which describes the service that is being served to the user. Specifies producer/administrator use-case.

The term Digital Forensic is set in parentheses, because it should be considered to remove all along, cf. the previous discussion.

6.2 Practical Considerations

In addition to theoretical implications, several practical considerations are proposed as measures to be considered when implementing and operating a DFaaS system.

There should be minimum competence criteria's

It is important that the users in a DFaaS system holds the appropriate training. It is outside the scope of this study to suggest exactly what this training should include, but the measures that was described in Section 5.5 gives some relevant headings: "The users should have the right *competence* and *comprehension* of how the functionality and results of the work methods in a DFaaS system plays out in the investigation, and they would need *capability* and *skills* to handle the application."

These are more specific to DFaaS. In addition, there should be general digital investigation competence criteria.

Education should be part of the system
By integrating education into the system, the path to become a user of the system with the required training becomes clear and defined. Time must be allowed to build competence. I would argue this is more important than the labor force itself.

The digital investigators need to be coupled with the detectives
There are different levels and areas of expertise in a DFaaS system, which is a very positive thing. By bringing the tactical knowledge together with the technical knowledge, the investigation capacity would be more resilient and flexible. Regardless, respecting that machines have faults, just like humans, for the level of trust on the data to be satisfiable, the digital expertise must be implemented into the process.

The importance of verification procedures
This research has led to the identification of several advantages for the investigation, as well as challenges where most of them are general to digital forensics but nevertheless applies here as well.

Digital Forensics as a Service could increase the efficiency of the investigation, especially in the processing phase. It may increase the accessibility of the data by giving investigators early access, and case investigators will get more involved in the digital investigation. It has the potential to enhance the investigative cooperation and focuses the expertise on tasks that requires such attention, by the release of such as administrative tasks.

It is on the other hand critical to have a supportive infrastructure, because performance issues and delays could have an opposite effect to the previous mentioned. By involving case investigators, which are often non-technical, more in the digital investigation could have some drawbacks as well.

The act of increasing user-friendliness and making the interface simpler, would often mean that data is abstracted to a higher level, which is usually good since this is generally more comprehensive for the human to understand. But there are some pitfalls to this by moving the investigator further away from the source/raw data, both in terms of the data itself (i.e., from binary- to readable data), but also physical distance (i.e., from server to client), as identified in Chapter 2.3.1 and 2.3.2. Risk factors such as an over-reliance on the tools, misinterpretations and tool bugs, are present in all digital forensics scenarios involving the investigation of digital information.

Several measures to deal with these kinds of challenges were identified in Chapter 5. With regards to the methods that could be used to mitigate for potential errors, the act of verification stands central to the digital investigator. This role is central to the quality control of the digital investigation, and therefore an idea for a concept that gives an example of how this could be carried out in an investigation is presented.

Practical example on verification procedures

- ✓ **Multiple sources verification: Data validation by verifying via multiple sources**

  Motivation

  Traditionally, the act of verifying everything could be a burden for the digital investigator. With an increase in cases with digital evidence the backlogs are increasing, analysis computers are struggling, management is demanding, and case detectives are crying for help. It could be a real challenge to prioritise checking everything multiple times via numerous sources, potentially doubling the investigation time because of ruthless attention to detail. However, it should be crucial, and a given.

  Digital Forensics as a Service releases the digital investigators from much of the overhead of tasks and provides for more time and methods/opportunities to verify. The case detectives become more involved in the digital investigation and can verify findings against tactical information and include more peer review, since the whole team can access the evidence.

  The concept called "Multiple sources verification" is inspired from dual tool verification (see Section 2.2.2 and 2.3.1) but is a more general approach. Casey listed different approaches to verification (36); hash comparison, comparing results of multiple tools, checking data at a low level and peer review. The framework has its basis in these methods.

  To give an example, the scenario utilised throughout the thesis will be concluded.

Scenario, part 4

Harry was investigating a homicide case. Since the trojan horse case he had learned from his earlier mistakes and had invested in training in digital forensics. The local Digital Forensics unit had also been reinforced as a result of an increased budget to the digital investigation field.

The computer of the deceased was found to be compromised and remotely controlled in an active RDP session. An IP from the remote session was traced to a person (hereby called the subject), which was apprehended. It was seized a computer, a 16-bay NAS and a mobile phone from the subject.

There was about 56 TB of data in the case, which was processed in the DFaaS system directly from the police server in Torskevik. The system had since last time been expanded with dedicated servers and system operators located in the police district. It was also dedicated fibre lines to the detectives and digital investigators offices.

A preliminary triage of the data revealed a large amount of video and image material showing surveillance of different people's homes. Data filtering techniques and video snapshots helped the investigators go through data very rapidly and it was revealed that several of the people in the footages were missing persons. It was also found chat messages on the computer and the phone of the subject, showing communication between the deceased and the subject, where the subject tries to warn the deceased that he is on a killing list. The strange thing was that duplicates of the messages were found, where one shows that the message is sent right after the murder, and the second shows that the message was sent 10 hours before. The device of the deceased that the messages was sent to, was unfortunately not found so it could not be correlated against the data from the subject.

The timestamps would be crucial to interpret correctly to see whether the subject warned the deceased. Was it before or after the murder? The local Digital Forensics unit assisted Harry and started by confirming the integrity of the images. The time zone settings of both the computer and the mobile phone was checked and it was discovered that there were discrepancies; the computer had UTC -8 and the phone had UTC +2. The local time zone was currently UTC +2.

The computer was investigated further, and it was found Unix valued timestamps in the databases of the application. The database was checked with two separate SQLite database viewers. The data was then controlled at a level deeper in the hexadecimal data, it was converted and verified that the Unix value was correct. The message sent time was showing 13:10.

The database of the phone was then checked with multiple tools and it was discovered data that suggested the message to having been sent 20:10. A second digital forensics expert confirmed the findings. They searched for settings or documentation for the application in what time zone the timestamps were stored in but found nothing.

The scenario was replicated by testing the application on similar devices, with an identical operating system and version of the app. The same test was also performed by the other expert.

The testing discovered that the computer application database stored the timestamps in local time, which was UTC -8 and the phone stored it as UTC 0. By converting the times to the actual time zone, it was concluded that the messages were sent 22:10 (UTC +2). Further investigation revealed that the subject changed the time zone of his computer right after the killing, but probably forgot to change it on the phone as well.

(The end)

Proposed framework

In Chapter 5, one of the suggestions was to integrate quality measures in the DFaaS system by making verification easier of sources and between tools, and also to have documentation as a feature in the application. According to the authors' experience, the police reports are generally not very detailed with regards to the documentation process and should arguably not be either. This report is a case document that is written for other investigators, attorneys, and court associates such as the judge and jury. It should have the most important conclusions and it should be easy to read.

The question is then, how are all the details with regards to the investigation and the analysis stored? A suggestion would be that this could be implemented in the DFaaS system, making it easy for the users to access, thereby acting as a natural part of the working process.

An example related to the scenario is given as an Excel diagram, shown in Appendix F: Multiple sources verification. The diagram seeks to integrate information from the verification process of traces/data/information and its intention is to support the report and documentation process by showing;

- Type of trace/data/information that are subject to verification,
- from what item and source,

- the tool or method used,
- in what abstraction format,
- what the hypothesized result was (remember to confirm/test this),
- if the result diverts from the majority,
- if it is peer reviewed,
- if the peer finds the result to deviate,
- if it has been tested,
- if the test result shows divergence,
- and comments to the verification procedure.

In the appendix, the most crucial stages of checking when the messages were sent is included. Tactical information taken from the statements of the parties (suspect/witness) could also be included here, and this way the tactical investigators also becomes involved in the process of verifying data and information.

Using Excel for this purpose of documenting the verification process is most certainly not the best method, but the intention here is just to provide an example of how the data or information could be verified and how it could be recorded in writing.

At last there is a conclusion based on the verification process, with the estimated level of certainty. "How sure are we that this information is correct?" This is based on the method of estimating and categorizing uncertainty in digital data from Casey: "In addition to using corroborating data from multiple, independent sources, forensic examiners should attempt to rate their level of confidence in the relevant digital evidence" (58, p. 41). He argues this would help the rest of the persons involved in the investigation to assess the reliability of the digital evidence: "In addition to providing forensic examiners with a practical method for estimating uncertainty, this heuristics approach allows investigators, attorneys, judges, and jurors who do not have a deep technical understanding of network technology to assess the reliability of a given piece of digital evidence" (58, p. 16). This grading was developed for networked systems, but it could be argued that it could be used for other type of information and data as well.

There are 7 levels of rating the certainty (0 – 6), with the following descriptions:

- C0: Erroneous/incorrect (evidence contradicts known facts).
- C1: Highly uncertain (missing entries or signs of tampering).
- C2: Somewhat uncertain (only single source of evidence).
- C3: Possible (inconsistencies between various sources).
- C4: Probable (multiple sources of evidence available, details can be corroborated).
- C5: Almost certain (complete agreement between sources).
- C6: Certain (complete agreement between tamper proof sources – not achievable according to Casey since no tamper proof digital evidence exists, which I would agree even 17 years after the article was written).

6.3 Further Research

The following subjects are proposed to study further based on the findings in this study.

Work towards finding a compatible process model for DFaaS
As was identified in Section 5.2, identifying the most fitting process model for DFaaS would require a better data foundation in terms of variety and quantity in tools, models and respondents. This was outside of the scope of this thesis, but it would be a valuable addition to how law enforcement agencies could best take advantage of DFaaS in accordance with the principles of the investigation.

This is a central proposal for further studies.

Investigate if DFaaS causes less or more errors in the investigation
By comparing with traditional digital investigation, it would give a good baseline on if the implementation of the system was ideal or not. Do the same errors appear in DFaaS as in the traditional system, or are there new ones?

Future studies should include more non-experienced users
With regards to the ability to generalise from this study and its external validity, the non-experienced users should have a greater part of the study to give a realistic baseline of the organization.

Conduct a case study of DFaaS over a longer period
This could prove valuable to get a more comprehensive view on the experiences and development of the system. In this study, a cross section of users at a certain time was chosen (cross-sectional study). The system could have had temporary issues recently[22], which would have affected the answers of the participants, but that does not mean it would be representative for when the system worked ideally for the rest of the year. By having conducted a longitudinal study, the data related to the characteristics under investigation are collected at various times. Such as temporary issues would then be easier to identify, and the results would not be volatile to the same degree.

Include topics left out of scope
The topics not included in the scope, such as; technical details, privacy, security, other system levels, other kinds of users in the investigation and whether the other phases (planning, incident response etc.) are compatible with DFaaS are relevant to include in other studies. This to get a more comprehensive view and identify other factors that is important to consider when implementing DFaaS.

Expand on the multiple sources' verification framework
The draft given in the previous chapter was just an example of how the verification procedures could be logged. Can this be programmed into an application, so it possibly could be implemented into something like DFaaS? Digital Forensics as a Service merges a lot of features ("all-in-one") that improves the usability. Quality procedures such as

[22] This was confirmed to be the case in this study, se strengths and weaknesses in the last chapter.

documenting circumstances of the analysis, not just the analysis itself should be integrated and easy to use, also for the case detectives.

Checklists are another great example of this; when the case detective logged into the system they will know what is verified and what is not. After the review of data is done, it could be sent to “technical review”, so that more competent personnel will address the findings and evaluations of hypotheses. This way peer review and technical review is a natural part of the investigation.

References

1. Garfinkel SL. Digital forensics research: The next 10 years. Digital Investigation. 2010;7:S64-S73.

2. Komplett.no consumer electronics web store [cited 07. apr 2019]. Available from: 1999: https://web.archive.org/web/19991105230113/http://www.komplett.no:80/waggon/Harddisk.html 2009: https://web.archive.org/web/20090629031441/http://www.komplett.no:80/k/kl.aspx?mfr=&bn=10098&sortBy=p&zbo2=true 2019: https://www.komplett.no/category/10088/datautstyr/lagring/harddisker/ssd?sort=Price%3AASCENDING&hits=120&nlevel=10000%C2%A728001%C2%A710088%C2%A710098.

3. Quick D. Big digital forensic data. : Data reduction framework and selective imaging : Volume 1. Singapore: Springer; 2018.

4. Zawoad S, Hasan R, editors. Digital Forensics in the Age of Big Data: Challenges, Approaches, and Opportunities. 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems; 2015 24-26 Aug. 2015.

5. Franke K, Srihari SN, editors. Computational Forensics: An Overview2008; Berlin, Heidelberg: Springer Berlin Heidelberg.

6. van Baar RB, van Beek HMA, van Eijk EJ. Digital Forensics as a Service: A game changer. Digital Investigation. 2014;11:S54-S62.

7. Kaufman J. The Personal MBA: Master the Art of Business: Penguin Publishing Group; 2010.

8. Kowalski S. IT Insecurity: A Multi-disciplinary Inquiry: Univ.; 1994.

9. Årnes A. Digital Forensics: John Wiley & Sons; 2017.

10. Andersen S. Technical Report: A preliminary Process Model for Investigation. SocArXiv. 2019.

11. Flamez B, Lenz AS, Balkin RS, Smith RL. The Literature Review. 2017:93-110.

12. Gartner IT Glossary Gartner [cited 05. Jan 2019]. Available from: https://www.gartner.com/it-glossary/big-data/.

13. Guarino A. Digital Forensics as a Big Data Challenge. In: Reimer H, Pohlmann N, Schneider W, editors. ISSE 2013 Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe 2013 Conference. Wiesbaden: Springer Fachmedien Wiesbaden; 2013. p. 197-203.

14. Du X, Le-Khac N-A, Scanlon M. Evaluation of Digital Forensic Process Models with Respect to Digital Forensics as a Service. 2017.

15. Lillis D, Becker B, O'Sullivan T, Scanlon M. Current challenges and future research areas for digital forensic investigation. arXiv preprint arXiv:03850. 2016.

16. Palmer G. A road map for digital forensic research2001. 27-30 p.

17. Guidance Encase web page [Available from: https://www.guidancesoftware.com/encase-forensic.

18. Magnet Axiom web page [Available from: https://www.magnetforensics.com/products/magnet-axiom/.

19. Cellebrite UFED web page [Available from: https://www.cellebrite.com/en/products/ufed-ultimate/.

20. Accessdata FTK web page [Available from: https://accessdata.com/products-services/forensic-toolkit-ftk.

21. Ayers D. A second generation computer forensic analysis system. Digital Investigation. 2009;6:S34-S42.

22. Casey E, Ferraro M, Nguyen L. Investigation delayed is justice denied: proposals for expediting forensic examinations of digital evidence. Journal of forensic sciences. 2009;54(6):1353-64.

23. Garfinkel S. Lessons learned writing digital forensics tools and managing a 30TB digital evidence corpus. Digital Investigation. 2012;9:S80-S9.

24. Clemens J. Automatic classification of object code using machine learning. Digital Investigation. 2015;14:S156-S62.

25. James JI, Gladyshev P. Challenges with automation in digital forensic investigations. arXiv:13034498v1 [csCY]. 2013.

26. Vincze EA. Challenges in digital forensics. Police Practice and Research. 2016;17(2):183-94.

27. The Attorney General in Norway. Quality requirements to the criminal proceedings in the police and at the public prosecutor. Circular letter nr. 3/2018.

28. Gogolin G. The Digital Crime Tsunami. Digital Investigation. 2010;7(1):3-8.

29. Al Fahdi M, Clarke NL, Li F, Furnell SM. A suspect-oriented intelligent and automated computer forensic analysis. Digital Investigation. 2016;18:65-76.

30. Breaking the backlog of digital forensic evidence 2013 [cited 06. jan 2019]. Available from: https://www.helpnetsecurity.com/2013/12/23/breaking-the-backlog-of-digital-forensic-evidence/.

31. van Beek HMA, van Eijk EJ, van Baar RB, Ugen M, Bodde JNC, Siemelink AJ. Digital forensics as a service: Game on. Digital Investigation. 2015;15:20-38.

32. STRASAK rapport 2015 [cited 06. jan 2019]. Available from: https://www.politiet.no/globalassets/04-aktuelt-tall-og-fakta/strasak/2015/strasak-3-2015.pdf. Norwegian.

33. STRASAK rapport 2017 [cited 06. jan 2019]. Available from: https://www.politiet.no/globalassets/04-aktuelt-tall-og-fakta/strasak/2017/strasak-2017.pdf. Norwegian.

34. STRASAK rapport 2018 [cited 06. jan 2019]. Available from: https://www.politiet.no/globalassets/04-aktuelt-tall-og-fakta/strasak/2018/201703842-1_v3_strasak-rapport-2.tert-2018---endelig-v2.pdf. Norwegian.

35. NOU: Digital sårbarhet - sikkert samfunn 2015 [cited 06. Jan 2019]. Available from: https://www.regjeringen.no/contentassets/fe88e9ea8a354bd1b63bc0022469f644/no/pdfs/nou201520150013000dddpdfs.pdf. Norwegian.

36. Casey E. Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. 3rd ed. ed: United States: Academic Press; 2011.

37. Shalaginov A, Johnsen JW, Franke K, editors. Cyber crime investigations in the era of big data. 2017 IEEE International Conference on Big Data (Big Data); 2017 11-14 Dec. 2017.

38. Carrier B, Spafford EH. . International Journal of digital evidence. 2003;2(2):1-20.

39. Rahayu S, Robiah Y, Sahib S. Mapping Process of Digital Forensic Investigation Framework2008.

40. Abulaish M, Haldar N. Advances in Digital Forensics Frameworks and Tools: A Comparative Insight and Ranking2018. 95-119 p.

41. Vidwarshi S, Chandra N. Analysis Of Development Phases In Digital Forensics Model. International Journal of Advance Computational Engineering and Networking (IJACEN). 2015;Volume-3(Issue-8):pp 89-95.

42. Kohn MD, Eloff MM, Eloff JHP. Integrated digital forensic process model. Computers & Security. 2013;38:103-15.

43. Baryamureeba V, Tushabe F, editors. The enhanced digital investigation process model. Proceedings of the Fourth Digital Forensic Research Workshop; 2004.

44. Beebe NL, Clark JG. A hierarchical, objectives-based framework for the digital investigations process. Digital Investigation. 2005;2(2):147-67.

45. Quick D. Big digital forensic data. : Data reduction framework and selective imaging : Volume 2. Singapore: Springer; 2018.

46. Casey E. Differentiating the phases of digital investigations. Digital Investigation. 2016;19:A1-A3.

47. Sunde N. Non-technical Sources of Errors When Handling Digital Evidence within a Criminal Investigaton 2017.

48. Watson DL, Jones A. Digital forensics processing and procedures: Meeting the requirements of ISO 17020, ISO 17025, ISO 27001 and best practice requirements: Newnes; 2013.

49. Garfinkel S, Farrell P, Roussev V, Dinolt G. Bringing science to digital forensics with standardized forensic corpora. Digital Investigation. 2009;6:S2-S11.

50. Brunty J. Validation of Forensic Tools and Software: A Quick Guide for the Digital Forensic Examiner2011.

51. Garrie DB. Digital forensic evidence in the courtroom: Understanding content and quality. Nw J Tech Intell Prop. 2014;12:i.

52. Homem I. Towards Automation in Digital Investigations : Seeking Efficiency in Digital Forensics in Mobile and Cloud Environments [Licentiate thesis, comprehensive summary]. Stockholm: Department of Computer and Systems Sciences, Stockholm University; 2016.

53. Hart SV, Ashcroft J, Daniels DJ. Forensic examination of digital evidence: a guide for law enforcement. National Institute of Justice NIJ-US, Washington DC, USA, Tech Rep NCJ. 2004;199408.

54. Williams J. ACPO Good Practice Guide for Digital Evidence. Metropolitan Police Service, Association of chief police officers, GB. 2012.

55. SWGDE establishing confidence in digital forensic results by error mitigation analysis, version: 2.0. Scientific Working Group on Digital Evidence. 2018.

56. Carrier B. Defining digital forensic examination and analysis tools using abstraction layers. International Journal of digital evidence. 2003;1(4):1-12.

57. Friheim I. Practical use of dual tool verification in computer forensics. Unpublished; 2016.

58. Casey E. Error, uncertainty, and loss in digital evidence. International Journal of Digital Evidence

2002;1(2):1-45.

59. Bhoedjang RA, van Ballegooij AR, van Beek HM, van Schie JC, Dillema FW, van Baar RB, et al. Engineering an online computer forensic service. Digital Investigation. 2012;9(2):96-108.

60. Chang WL, Roy A, Grady N, Reinsch R, Underwood M, Fox G, et al. NIST big data interoperability framework. 2018.

61. Dykstra J, Sherman AT. Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. Digital Investigation. 2012;9:S90-S8.

62. Techopedia.com. What is Automation? - Definition from Techopedia. [cited 17. feb 2019]. Available from: https://www.techopedia.com/definition/32099/automation

63. Deloitte. Policing 4.0 - Deciding the future of policing in the UK. 2018.

64. Beebe N, editor Digital forensic research: The good, the bad and the unaddressed. IFIP International Conference on Digital Forensics; 2009: Springer.

65. Kovar D. Intergriography: A Journal of Broken Locks, Ethics, and Computer Forensics 2009. [cited 19. feb 2019]. Available from: https://integriography.wordpress.com/2009/11/17/the-value-of-push-button-forensics/.

66. Casey E. Cutting corners: Trading justice for cost savings. Digital investigation. 2006;4(3):185-6.

67. Rogers MK, Goldman J, Mislan R, Wedge T, Debrota SJJoDF, Security, Law. Computer forensics field triage process model. 2006;1(2):2.

68. Mislan RP, Casey E, Kessler GCJDI. The growing need for on-scene triage of mobile devices. 2010;6(3-4):112-24.

69. Moser A, Cohen MI. Hunting in the enterprise: Forensic triage and incident response. Digital Investigation. 2013;10(2):89-98.

70. Koopmans MB, James JI. Automated network triage. Digital Investigation. 2013;10(2):129-37.

71. Reyes A, O'Shea K, Steele J, Hansen J, Jean B, Ralph T. Digital forensics and analyzing data, cyber crime investigations. Syngress, Burlington. 2007:219-59.

72. Roussev V, Quates C. Content triage with similarity digests: The M57 case study. Digital Investigation. 2012;9:S60-S8.

73. Scanlon M, Kechadi M-T, editors. Online acquisition of digital forensic evidence. International Conference on Digital Forensics and Cyber Crime; 2009: Springer.

74. Watkins K, McWhorte M, Long J, Hill B. Teleporter: An analytically and forensically sound duplicate transfer system. Digital Investigation. 2009;6:S43-S7.

75. Garfinkel SL. Forensic feature extraction and cross-drive analysis. Digital Investigation. 2006;3:71-81.

76. Richard III GG, Grier J. Rapid forensic acquisition of large media with sifting collectors. 2015.

77. Greiner L. Sniper forensics. Networker. 2009;13(4):8-10.

78. Alink W, Bhoedjang RAF, Boncz PA, de Vries AP. XIRAF – XML-based indexing and querying for digital forensics. Digital Investigation. 2006;3:50-8.

79. Beebe N, Clark J, editors. Dealing with terabyte data sets in digital investigations. IFIP International Conference on Digital Forensics; 2005: Springer.

80. Brown R, Pham B, de Vel O, editors. Design of a digital forensics image mining system. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems; 2005: Springer.

81. Huang J, Yasinsac A, Hayes PJ, editors. Knowledge Sharing and Reuse in Digital Forensics. 2010 Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering; 2010 20-20 May 2010.

82. Roussev V, Richard III GG, editors. Breaking the performance wall: The case for distributed digital forensics. Proceedings of the 2004 digital forensics research workshop; 2004.

83. Pringle N, Burgess M. Information assurance in a distributed forensic cluster. Digital Investigation. 2014;11:S36-S44.

84. Richard III GG, Roussev V. Next-generation digital forensics. Communications of the ACM. 2006;49(2):76-80.

85. Cruz F, Moser A, Cohen M. A scalable file based data store for forensic analysis. Digital Investigation. 2015;12:S90-S101.

86. Stelly C, Roussev V. SCARF: A container-based approach to cloud-scale digital forensic processing. Digital Investigation. 2017;22:S39-S47.

87. Carrier BD. Volume analysis of disk spanning logical volumes. Digital Investigation. 2005;2(2):78-88.

88. Raghavan S, Raghavan SV, editors. AssocGEN: Engine for analyzing metadata based associations in digital evidence. 2013 8th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE); 2013 21-22 Nov. 2013.

89. Case A, Cristina A, Marziale L, Richard GG, Roussev V. FACE: Automated digital evidence discovery and correlation. Digital Investigation. 2008;5:S65-S75.

90. Raghavan S, Clark A, Mohay G, editors. FIA: an open forensic integration architecture for composing digital evidence. International Conference on Forensics in Telecommunications, Information, and Multimedia; 2009: Springer.

91. Cohen MI. PyFlag – An advanced network forensic framework. Digital Investigation. 2008;5:S112-S20.

92. Hoelz BW, Ralha CG, Geeverghese R, editors. Artificial intelligence applied to computer forensics. Proceedings of the 2009 ACM symposium on Applied Computing; 2009: ACM.

93. Karabiyik U. Building an intelligent assistant for digital forensics. 2015.

94. Irons A, Lallie H. Digital forensics to intelligent forensics. Future Internet

2014;6(3):584-96.

95. Mukkamala S, Sung AH. Identifying significant features for network forensic analysis using artificial intelligent techniques. International Journal of digital evidence. 2003;1(4):1-17.

96. Adedayo OM. Big data and digital forensics. 2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF). 2016:1-7.

97. Beebe NL, Clark JG. Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results. Digital investigation. 2007;4:49-54.

98. Raghavan S. Digital forensic research: current state of the art. CSI Transactions on ICT

2013;1(1):91-114.

99. Raghavan S, Raghavan SV, editors. A study of forensic & analysis tools. 2013 8th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE); 2013 21-22 Nov. 2013.

100. Wen Y, Man X, Le K, Shi W, editors. Forensics-as-a-service (faas): computer forensic workflow management and processing using cloud. The Fifth International Conferences on Pervasive Patterns and Applications; 2013: Citeseer.

101. Mell PM, Grance T. SP 800-145. The NIST Definition of Cloud Computing. National Institute of Standards \& Technology; 2011.

102. Ruan K, Carthy J, Kechadi T, Crosbie M, editors. Cloud forensics: An overview. proceedings of the 7th IFIP International Conference on Digital Forensics; 2011.

103. Lee J, Hong D, editors. Pervasive forensic analysis based on mobile cloud computing. Multimedia Information Networking and Security (MINES), 2011 Third International Conference on; 2011: IEEE.

104. Lee J, Un S, editors. Digital forensics as a service: A case study of forensic indexed search. 2012 International Conference on ICT Convergence (ICTC); 2012 15-17 Oct. 2012.

105. Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Communications of the ACM. 2008;51(1):107-13.

106. Roussev V, Wang L, Richard G, Marziale L, editors. A cloud computing platform for large-scale forensic computing. IFIP International Conference on Digital Forensics; 2009: Springer.

107. Shende JRG. Forensics as a Service. Cybercrime and Cloud Forensics: Applications for Investigation Processes: IGI Global; 2013. p. 266-90.

108. Didone D, de Queiroz RJ, editors. Forensic as a service-FaaS. Proceedings of the Sixth International Conference on Forensic Computer Science (ICoFCS); 2011.

109. Marturana F, Tacconi S, Italiano GF. A Forensic-as-a-Service Delivery Platform for Law Enforcement Agencies. 42013. p. 313-30.

110. Pichan A, Lazarescu M, Soh ST. Cloud forensics: Technical challenges, solutions and comparative analysis. Digital Investigation. 2015;13:38-57.

111. Zeng G. Research on Digital Forensics Based on Private Cloud Computing. 2014;2:24-9.

112. van Beek HMA. Digital Forensics as a Service: an update. Digital Forensics Research Workshop; Seattle, WA, USA. 2016.

113. Akbari H, Land F. Socio-technical theory IS Research Wiki. 2010 [updated 2016 20. mar; cited 03. jan 2019]. Available from: https://is.theorizeit.org/wiki/Socio-technical_theory.

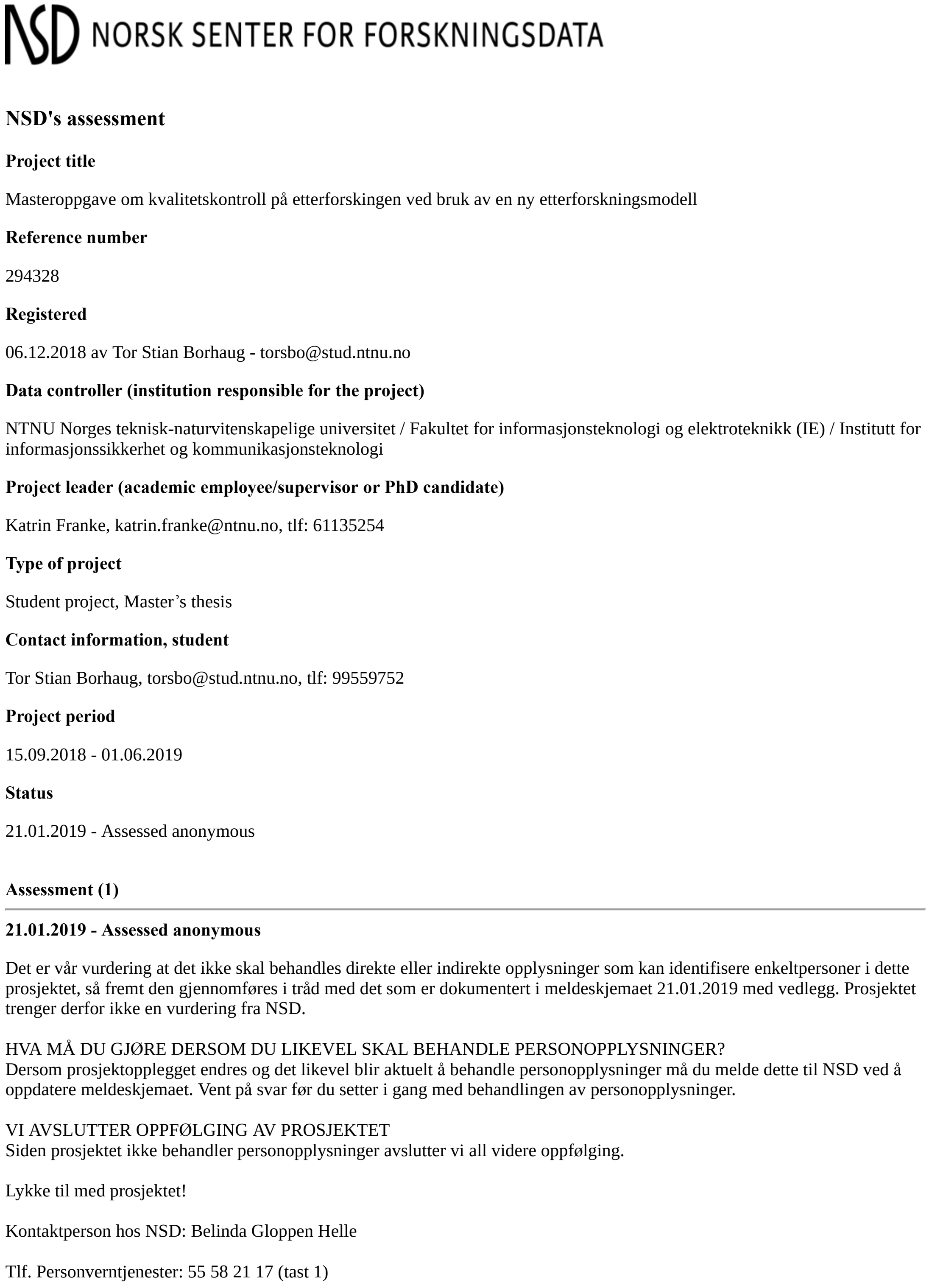
114. Leedy PD, Ormrod JE. Practical research : planning and design. 11th ed. ed. Boston: Pearson; 2015.

115. Kripos (NCIS): Teknisk utprøving av Hansken (Technical review of Hansken) - Exempt from public disclosure. 2018.

116. Creswell JW, Poth CN. Qualitative inquiry and research design: Choosing among five approaches: Sage publications; 2017.

117. Kim DH. Introduction to systems thinking: Pegasus Communications Waltham, MA; 1999.

Appendix A: Approval from the NSD



NSD NORSK SENTER FOR FORSKNINGSDATA

NSD's assessment

Project title

Masteroppgave om kvalitetskontroll på etterforskingen ved bruk av en ny etterforskningsmodell

Reference number

294328

Registered

06.12.2018 av Tor Stian Borhaug - torsbo@stud.ntnu.no

Data controller (institution responsible for the project)

NTNU Norges teknisk-naturvitenskapelige universitet / Fakultet for informasjonsteknologi og elektroteknikk (IE) / Institutt for informasjonssikkerhet og kommunikasjonsteknologi

Project leader (academic employee/supervisor or PhD candidate)

Katrin Franke, katrin.franke@ntnu.no, tlf: 61135254

Type of project

Student project, Master's thesis

Contact information, student

Tor Stian Borhaug, torsbo@stud.ntnu.no, tlf: 99559752

Project period

15.09.2018 - 01.06.2019

Status

21.01.2019 - Assessed anonymous

Assessment (1)

21.01.2019 - Assessed anonymous

Det er vår vurdering at det ikke skal behandles direkte eller indirekte opplysninger som kan identifisere enkeltpersoner i dette prosjektet, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet 21.01.2019 med vedlegg. Prosjektet trenger derfor ikke en vurdering fra NSD.

HVA MÅ DU GJØRE DERSOM DU LIKEVEL SKAL BEHANDLE PERSONOPPLYSNINGER?
Dersom prosjektopplegget endres og det likevel blir aktuelt å behandle personopplysninger må du melde dette til NSD ved å oppdatere meldeskjemaet. Vent på svar før du setter i gang med behandlingen av personopplysninger.

VI AVSLUTTER OPPFØLGING AV PROSJEKTET
Siden prosjektet ikke behandler personopplysninger avslutter vi all videre oppfølging.

Lykke til med prosjektet!

Kontaktperson hos NSD: Belinda Gloppen Helle

Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix B: Information letter

Are you interested in taking part in the research project:
"The paradox of automation with regards to
Digital Forensics as a Service"?

This is an inquiry about participation in a research project where the main purpose is to gather data regarding experiences with the use of the "Digital Forensics as a Service" model **Hansken**. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

The purpose of the study is to investigate how the digital investigator can maintain the quality on the investigation with this model. The paradox of automation can be defined as: The more efficient the automated system, the more crucial the human contribution of the operators. Humans are less involved, but their involvement becomes more critical. Digital Forensics as a Service (DFaaS) is a model that automates parts of the processing and investigation of cases regarding for example big data.

Police officer and MSc student

I work as a digital forensics investigator at the NCIS (National Criminal Investigation Service) in Norway, and this work is part of my master's thesis at NTNU (Norwegian University of Science and Technology). I want to find out if it is easier for the digital investigator to maintain the quality of the analysis in the DFaaS model compared to the traditional model used in digital forensics. Your contribution will be used as a data basis to do a risk-based analysis of DFaaS.

Why are you being asked to participate?

The target group for this survey are users/administrators of Hansken in the Dutch law enforcement.

What does participation involve for you?

If you chose to take part in the project, this will involve that you fill in an online survey. It will take approx. 10 minutes. The survey includes questions about your experience with Hansken and how it has affected your professional role. Participation in the study is voluntary and your identity will be hidden. When hidden identity is used in surveys, no identifiable information, such as browser type and version, internet IP address, operating system, or e-mail address, will be stored with the answer. This is to protect the respondent's identity.

NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?

If you have any questions about the project, please contact Tor Stian Borhaug, MSc student at torsbo@stud.ntnu.no or Katrin Franke, Professor in Informatics at katrin.franke@ntnu.no

Yours sincerely,

Tor Stian Borhaug

Appendix C: Survey of Dutch police

Survey on Hansken and Digital Forensics as a Service (Dutch police)

The following survey will ask you some questions about your experience with Hansken/Xiraf in regard to it being a Digital Forensics as a Service tool. The survey focuses on the usage of the tool in investigations, if it have affected your role and if it is compatible with the digital forensics process model.

The survey will take about 10 minutes to complete. If you don't have an answer, just leave it open, that's entirely up to you. Please do NOT use any names or other information that could identify any persons, because the survey is anonymous.

I really appreciate you taking the time, thank you.

Your identity will be hidden.

When hidden identity is used in surveys, no identifiable information, such as browser type and version, internet IP address, operating system, or e-mail address, will be stored with the answer. This is to protect the respondent’s identity.

1. Gender

Name	Percent
Male	77.8%
Female	22.2%
N	27

2. What role best describe your current position?

Name	Percent
Digital forensics investigator	48.1%
Cybercrime/specialist investigator	7.4%
Technical system operator	0.0%
Tactical investigator	18.5%
Other investigator	0.0%
Operative police	0.0%
Engineer/IT	0.0%
Administration	3.7%
Supervisor/chief/manager	3.7%
Lawyer/attorney	0.0%
Other	18.5%
N	27

3. Have you completed a police education?

Name	Percent
Yes	74.1%
No	25.9%

N	27

4. For how long have you been employed in law enforcement?

Name	Percent
0-3 years	18.5%
4-7 years	7.4%
8-12 years	3.7%
13+ years	63.0%
Other	7.4%
N	27

5. Where do you work?

Name	Percent
National police unit	37.0%
Regional police unit	40.7%
Local police unit	0.0%
NFI	3.7%
Other	18.5%
N	27

6. Is digital investigation your occupation and do you give guidance/support to less technical experienced investigators, or/and do you verify their work?

Name	Percent
Yes and I give guidance, support and verifies their findings	33.3%
Yes. I give guidance and support, but I do not verify their findings	37.0%
No	18.5%
Other	11.1%
N	27

7. Do you use Hansken in your work?

Name	Percent
Yes	51.9%
Not now, but previously	25.9%
I have only used Xiraf and not Hansken	7.4%
No, never have	14.8%
N	27

8. When did you use it?

- At a large case where Hansken was the only way to process the data via scripts.
- several years ago
- From 2013-2017. Used the back-end. We did have or OWN Xiraf server
- During an investigation in 2015
- 2014 – 2018
- 2015-2017
- from the start until 2 yrs ago
- A few months ago
- 2016

9. What is the main reason you have not used Hansken/Xiraf?

Name	Percent
The place I work at do not use it	50.0%
It is available, but I use other tools	0.0%
I don't know	25.0%
Other	25.0%
N	4

10. How much experience do you have with Hansken/Xiraf?

Name	Percent
Less than 1 year	17.4%
1-2 years	43.5%
3-4 years	13.0%
5-6 years	13.0%
7-9 years	8.7%
Other	4.3%
N	23

11. How often do you use/did you use Hansken/Xiraf?

Name	Percent
Every day	0.0%
Almost daily	13.0%
A couple of times per week	8.7%
Weekly	13.0%
Couple of times per month	13.0%
Monthly or less	34.8%
Not using it for the moment	0.0%
Other	17.4%
N	23

12. Do you use only Hansken/did you only use Xiraf when working with digital evidence?

Name	Percent
Yes	13.0%
No	87.0%
N	23

13. What other tools do you use?

Name	Percent
X-Ways	15.0%
Encase	80.0%
FTK	75.0%
Axiom/IEF	75.0%
Belkasoft Evidence center	10.0%
Netclean/Griffeye	0.0%
Sleuthkit/Autopsy	40.0%
Cellebrite UFED	95.0%
Msab XRY	65.0%
OSForensics	10.0%
Linux distros (Ubuntu, SIFT etc)	70.0%
Other	20.0%
N	20

14. And why do you use other tools?

Name	Percent
To verify data/results	60.0%
Lack support in Hansken	45.0%
Have support, but not very good implemented in Hansken	15.0%
I am more used to use another tool for the task	35.0%
To get a better GUI	30.0%
Because I do not trust Hansken	5.0%
I don't know	0.0%
Other	35.0%
N	20

15. On a scale from 1 to 5 how much do you trust Hansken to give you valid data?

Name	Percent
1 (I trust it enough to only use this tool and confidently present my findings in the court)	8.7%
2 (I trust it, but I regularly use a secondary tool to verify data)	65.2%
3 (Neither trust or distrust it, results need to be verified and validated any way)	21.7%
4 (I do not trust it before the data gets verified and validated in multiple tools)	0.0%
5 (Do not trust it at all)	0.0%
Only used Xiraf	4.3%
I don't know	0.0%
Other	0.0%
N	23

16. On a scale from 1 to 5 how satisfied are you with how Hansken can be used to work with digital evidence?

Name	Percent
1 (It is absolute key to work with digital evidence, there are no better solutions that I know of)	4.5%
2 (It is a good tool to work with digital evidence)	54.5%
3 (It is OK, nothing more or less)	22.7%
4 (It is below average)	13.6%
5 (It is totally useless for the task)	0.0%
Only used Xiraf	4.5%
I don't know	0.0%
Other	0.0%
N	22

17. Any comments regarding validity of data in Hansken?

- Checksums of entire images to validate/check uploads.
- I fully trust it. The NFI stands for the forensic quality and I know them very well so I am confident Hansken is up to its task.
- sometimes unclear what (files, evidence, traces) Hansken supports and what not.
- I work with a lot of old files, which will be scanned to import it into Hansken. Too often the files are not fully readable. And salso, sometimes people make writing errors, especially in names.
- It would be much better of it was easier to search with wildcards, or the searchprogramme should give suggestions if there are no exact matches with the search terms. Nowthe risk to miss something is pretty high.
- Not enough experience with Hansken to comment on this

- Xiraf did request to set the correct time zone by the investigator at the start of the investigation. This was a tricky malfunction for the integrity of the findings
- It is a tool used by detectives not allways by digital forensic investigators / specialists
- No

18. Any comments regarding satisfaction with the usage of Hansken?

- Better accessability to the Python API.
- Within Dutch Police, the Hansken GUI is not reachable from standard desktops. That is not Hansken's fault, but still needs to be fixed.
- The main reason I am such a Hansken fan, is it's ability to act as an enging in the forensic process. That's still something that is not very widespread, unfortunately.
- Slow performance. Instabale. Good software voor first review of evidence.
- great all-in-one tool , search options better than other programs (search across evidence, very limited in ftk/ufed)
- It's easy.
- Sometimes it takes zo much time to load and then you get an error code, like "An error occurred.."
- It is not really clear what the consequences of these errors are, of why they occur (should I restart? r can I just continue the searches?)
- I dont like it: the error messages, the report function, the limited functionalities, the way it looks. I'd rather use Axiom ten times thanu struggle with Hansken (i don't like the name either :-))
- Technical difficulties and performance are sometimes an issue
- Xiraf front and back end are made nu programmeurs and not by people who think about smart interfaces.
- The interface is confussing for a tactical investigator.
- I work i the Netherlands with Hanken. unfortunately the implementation of the tool was bad from the beginning en still is.
- The tool was ment for every Policeofficer in the field so they could use hansken from their policeaccount and look through the data.
- The Tool was very slow from the beginning. There was no storage for the big amount of data en the bandwith was en is very poor.
- So most of the time when i search for something i can go get a cup of coffee before i see any results. Even the team that is responisble for Hansken send a mail a week ago with apologies for the bad performace and failure of implementation of the tool.
- My meaning is thad when the bandwith wis good and storage too than Hansken is a great tool for the tactical investigator to work with. When that problem is not solved A tool like Axiom is much better for them.
- I do not really need hansken. There are better tools for specific jobs. I just want to be able to choose a tool that suits the job best. That means that I often use several tools in one case.
- The best software is developed by companies where software development is their core business. NFI has another core business...
- The interpretation of data is misleading for the task. It also takes quite some time to get up and running in Hansken. Totally not suited for in incident response cases.

19. Have you presented results from Hansken in the court?

Name	Percent
Yes	21.7%
No	73.9%
Only used Xiraf	4.3%
N	23

20. When in court, where you asked something about Hansken or how you came to your conclusions/results? (Check all the correct statements)

Name	Percent
Yes, they asked me how Hansken worked	0.0%

Yes, they asked me how I came to the conclusions/results	20.0%
Yes, they asked if the data was verified	0.0%
Yes, they asked if the data was reliable	0.0%
Yes, they asked if the data was verified in multiple tools	0.0%
Yes, they asked if the source data hashes was compared to the data in the presentation report	0.0%
We (police/prosecutor) had to bring in an expert witness to explain how Hansken works	0.0%
The defense had to bring in an expert witness to question Hansken	40.0%
Nothing about Hansken in particular	40.0%
Nothing about how I came to my conclusions	40.0%
Other	0.0%
N	5

21. Did you work with digital evidence before using Hansken/Xiraf?

Name	Percent
Yes	91.3%
No	8.7%
N	23

22. On a scale from 1 to 5 how have Hansken affected the effectiveness of your work process?

Name	Percent
1 (Much more effective)	0.0%
2 (More effective)	47.6%
3 (About the same as before using Hansken)	19.0%
4 (Less effective)	9.5%
5 (Much less effective)	9.5%
Only used Xiraf	0.0%
I don't know	14.3%
N	21

23. On a scale from 1 to 5 how have Hansken affected the quality of your work? (1=Much better quality/3=About the same/5=Made the quality worse)

Name	Percent
1 (Much better quality)	0.0%
2 (Better quality)	23.8%
3 (About the same as before using Hansken)	52.4%
4 ((Made the quality worse)	0.0%
5 (Made the quality much worse)	4.8%
Only used Xiraf	0.0%
I don't know	19.0%
N	21

24. On a scale from 1 to 5 how have Hansken affected the collaboration on cases? (1=Much better collaboration/3=About the same/5=Made the collaboration worse)

Name	Percent
1 (Much better collaboration)	9.5%
2 (Better collaboration)	52.4%
3 (About the same as before using Hansken)	23.8%
4 (Made the collaboration worse)	4.8%
5 (Made the collaboration much worse)	0.0%
Only used Xiraf	0.0%
I don't know	9.5%
N	21

25. Any comments to the effectiveness, quality or collaboration?

- Collaboration and sharing of (parts of) data is much easier with Hansken.
- The software is less effective because the upload of data to Hansken is very slow and unstabale.
- The big risk is that non-technical investigator make there own conclusions. The official result to be presented to court should always have a technical check
- It doesn’t always work well in international cases
- The UI needs improvement to be able to rate better on above questions
- no

26. How often are your findings verified by a digital forensics investigator?

Name	Percent
Every time	25.0%
Every other time	0.0%
About 50 % of the time	25.0%
Not very often	0.0%
Never	0.0%
Please comment why/why not	50.0%
N	4

27. How often do you follow up the work of the investigators going through data in Hansken?

Name	Percent
Every time	11.8%
Every other time	5.9%
About 50 % of the time	17.6%
Not very often	35.3%
Never	5.9%
Does not apply	23.5%
N	17

28. What do you follow up?

Name	Percent
General quality control	41.7%
Responding to technical questions	100.0%
Dual tool verification	16.7%
Check metadata, timestamps etc	75.0%
Other	8.3%
N	12

29. Why do you not follow up?

Name	Percent
Not part of my profession	60.0%
I do not need to	20.0%
I do not have the time	20.0%
I do not have knowledge of how to do so	0.0%
I don't know	0.0%
Other	20.0%
N	5

30. Have Hansken freed up your time to do more research, develop methods/scripts, dig deeper in analysis etc

Name	Percent
Yes	35.3%
No	58.8%
Does not apply (only used Hansken)	0.0%
I don't know	5.9%
N	17

31. Here are some statements on the use of Hansken, please rate them accordingly

Question	Average	N
In Hansken we get access to the data early in the case.	2.52	21
Hansken gives us complete data (e.g. deleted/unallocated, hidden, obfuscated data).	2.05	21
In Hansken the amount of "data noise" is reduced (in other words the data are to a degree relevant and we do not have to go through system files etc)	2.21	21
It is easy to identify a possible incident to be investigated in Hansken.	2.24	20
The data is classified correctly (similar data grouped together).	1.90	21
The data is well organized (e.g. the program will work for a child explotation case as well as a case regarding narcotics).	2.05	21
In Hansken we can strengthen/weaken hypoteses quickly.	2.05	21
It is easy to attribute evidences found to a specific user in Hansken.	1.79	21
It is easy to evaluate the findings in Hansken (e.g. to see if it is important data for the case).	1.95	20
The findings in Hansken are easy to interpret and explain for the investigation group etc. In other words it makes communication easier.	2.15	21
With Hansken we can recontruct the data we find (e.g. we can make test-cases to try to reproduce the data).	2.13	21
It is easy to review the data in Hansken after we are finished investigating/analysing it.	2.05	21
Hansken makes it easier to manage a case (have a good overview of the process).	2.11	21
It is possible to verify and validate the quality/correctness of data through the whole process in Hansken (from it is seized till it is presented in court).	2.21	21

32. In Hansken we get access to the data early in the case.

Name	Percent
Absolutely	19.0%
To some extent	38.1%
Not so much	14.3%
No	28.6%
Don´t know	0.0%
N	21

Comments

- it's a matter of work process. The Hansken work process is not uniformly defined
- Because of slow upload and unstable software

33. Hansken gives us complete data (e.g. deleted/unallocated, hidden, obfuscated data).

Name	Percent

Absolutely	28.6%
To some extent	47.6%
Not so much	4.8%
No	14.3%
Don´t know	4.8%
N	21

Comments

- I cannot judge the level completeness. Additional tools should be used to do that. Again, a matter of defining and following the work process
- Should be more easy to know what is what, especially in cases where lots of devices are being downloaded into Hansken

34. In Hansken the amount of "data noise" is reduced (in other words the data are to a degree relevant and we do not have to go through system files etc) .

Name	Percent
Absolutely	9.5%
To some extent	57.1%
Not so much	19.0%
No	4.8%
Don´t know	9.5%
N	21

Comments

- ou can filter away system files
- system files are not always easy to discern for a tactical user

35. It is easy to identify a possible incident to be investigated in Hansken.

Name	Percent
Absolutely	10.0%
To some extent	50.0%
Not so much	20.0%
No	5.0%
Don´t know	15.0%
N	20

Comments

- the timeline functionality helps

36. The data is classified correctly (similar data grouped together).

Name	Percent
Absolutely	28.6%
To some extent	52.4%
Not so much	9.5%
No	4.8%
Don´t know	4.8%
N	21

37. The data is well organized (e.g. the program will work for a child explotation case as well as a case regarding narcotics).

Name	Percent
Absolutely	23.8%
To some extent	52.4%
Not so much	9.5%
No	9.5%
Don´t know	4.8%
N	21

38. In Hansken we can strengthen/weaken hypoteses quickly.

Name	Percent
Absolutely	23.8%
To some extent	42.9%
Not so much	19.0%
No	4.8%
Don´t know	9.5%
N	21

Comments

- You never know what is missing (maybe we don't have all the digital information a suspect has saved); it is still just a tool and not absolute (especially since I work with old cases)

39. It is easy to attribute evidences found to a specific user in Hansken.

Name	Percent
Absolutely	33.3%
To some extent	47.6%
Not so much	4.8%
No	4.8%
Don´t know	9.5%
N	21

40. It is easy to evaluate the findings in Hansken (e.g. to see if it is important data for the case).

Name	Percent
Absolutely	20.0%
To some extent	65.0%
Not so much	5.0%
No	5.0%
Don´t know	5.0%
N	20

41. The findings in Hansken are easy to interpret and explain for the investigation group etc. In other words it makes communication easier.

Name	Percent
Absolutely	19.0%
To some extent	52.4%
Not so much	14.3%
No	9.5%
Don´t know	4.8%
N	21

42. With Hansken we can reconstruct the data we find (e.g. we can make test-cases to try to reproduce the data).

Name	Percent
Absolutely	23.8%
To some extent	28.6%

Not so much	4.8%
No	14.3%
Don´t know	28.6%
N	21

Comments

- not sure what you mean

43. It is easy to review the data in Hansken after we are finished investigating/analysing it.

Name	Percent
Absolutely	23.8%
To some extent	47.6%
Not so much	19.0%
No	4.8%
Don´t know	4.8%
N	21

Comments

- If you print it out or save the files you downloaded/ reviewed.

44. Hansken makes it easier to manage a case (have a good overview of the process).

Name	Percent
Absolutely	23.8%
To some extent	33.3%
Not so much	23.8%
No	4.8%
Don´t know	14.3%
N	21

Comments

- we mainly use it for storage and analysis, Hansken tells you nothing abou the further process of a case

45. It is possible to verify and validate the quality/correctness of data through the whole process in Hansken (from it is seized till it is presented in court).

Name	Percent
Absolutely	23.8%
To some extent	33.3%
Not so much	23.8%
No	9.5%
Don´t know	9.5%
N	21

Comments

- validating uploads is difficult without a hash for the entire image

46. Any final comments or things you want to add?

- "It is easy to identify a possible incident to be investigated in Hansken." - nice question, since it brings up the discussion of 'is Hansken an incident alerting system?'. In some ways, it should be, I think.
- "The data is well organized (e.g. the program will work for a child explotation case as well as a case regarding narcotics)." - the data is well organized. There are programs specialized in child exploitation investigations. Hansken is not a solution for everything, but it does give the investigator a great overview on the case, andis a great tool for zooming in on specific traces.
- Hansken's team at the NFI is always very good in communication and also very accessible for us at the National Police.
- That makes it easy to tackle problems and issues.
- user interface is versatile, but not always clear where your are or what more there is (e.g. going back to a list of keywords found, might re-order it)
- Hansken seems to be developed for investigators that don’t how to operate EnCase or FTK: beginners. The problem is that they don’t know how to explain what they see.
- After Xiraf we did start to use Intella. We had better results with the last tool. It processen the data better. Also the front-end was more smart.
- It is one of the few tools to handle lots of data at once.

Appendix D: Interview with NFI

(NFI´s answers in *italics)*

1. Approximately how many people are using Hansken with regards to digital investigations today in the Netherlands?

    - *Not answered*

2. How is the approx. split between the type of users of Hansken in a digital investigation?

	Less than 10%	10-30%	30-50%	50-70%	70-100%	I don't know
IT administrators/developers						
System operators						
Digital investigators						
Tactical investigators						

Any groups I forgot to mention?

- *Not answered*

3. Which group do you think should be able to understand the technical parts of the system, and at what level?

	No need to understand the technicalities	Should have a basic understanding	Should have a well understanding	Should know every detail	I don't know
Case investigators (less technical competent)	*X*				
Digital investigators (more technical competent)		*X*			
Administrators (administrating the system)			*X*		
Developers				*X*	

Any groups I forgot to mention?

- *Not answered*

4. I understand that training on Hansken is optional, but how is the training organized?
    - *Police academy*

5. I understand that Hansken provides extraction reports that help in finding out if an extraction passed (so as to verify the data). Is it possible to have a copy or example of what such a report contains?
    - *hansken support?*

6. If there are errors in the data (traces/results), will it be easy to catch? If so, is that possible for the investigator to do, or is it critical that a digital investigator that have more knowledge verifies the data, or will it be up to the system operator?
    - *Both. The Hansken case operator(s) are required to have a basic understanding what types of traces an extraction yields, and there are checks in place to spot the obvious errors (this usually focuses on verifying if the image yields a reasonable result set, for example if a lot of mailboxes are present and there are hardly any emails found, there might be a misinterpretation). There is also a mechanism to view how many errors occur in an extraction. A digital investigator is always required to verify the obtained result as it needs to be validated if the conclusions are sound.*

7. Are there any data/studies on the possible reduction of backlogs/investigation time due to the introduction of DFaaS? E.g. backlogs/cases before/after DFaaS.
    - *Difficult to measure, the police should know whether they should have done a case manually, and how much time it would take.*

8. If a police organization is to implement DFaaS, what is required in terms of numbers of personnel? Will it require more or less personnel than in a traditional digital investigation structure?
    - *The intention is that the number of digi's can shrink and become more specialists, instead of the spindle in all investigations. Perhaps simple things can be done more easily with minimal intervention of a digi. Based on my feelings, I say a reduction in staff because much more data is central.*

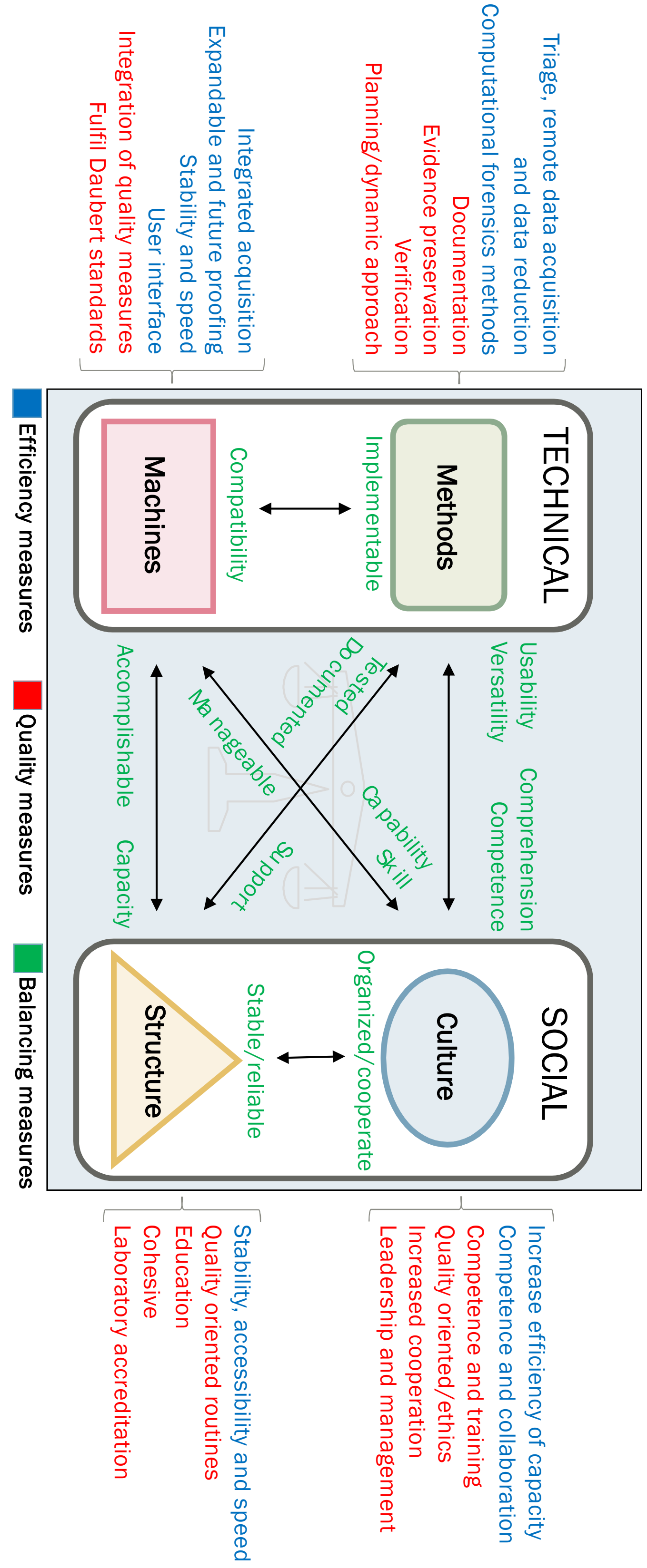
9. How can DFaaS be a solution to the challenges the police face in the big data development do you think? (e.g. faster processing, less "noise" in the data etc.)
    - *Not answered*

10. Try to list 3 possibilities that is introduced when implementing DFaaS (e.g. faster investigation, better collaboration etc.)
    - *1. Faster research, 2. Bundling of knowledge, 3. Bundling of cooperation*

11. Try to list 3 challenges that is introduced when implementing DFaaS (e.g. less human control over the data, technical investigators loose knowledge due to less "drilling" on manual tasks etc.)
    - *Not answered*

12. Will it be easier for the digital forensics investigator to maintain the quality of the digital investigation in a DFaaS based model versus in the traditional model? If so, why?
    - *Not answered*

Appendix E: Socio-technical balance in DFaaS



Appendix F: Multiple sources verification

Type of trace/data/information	Item	Source(s)	Tool/method	Abstraction/format	Hypothesized result	Divergent result	Peer reviewed	Peer divergence	Tested	Test divergence	Comment
Image integrity	A-1	Imagefile	DFaaS tool	Checksum/SHA256	OK		X				
	A-1	Imagefile	Computer forensic tool	Checksum/SHA256	OK						
	A-2	Imagefile	Phone forensic tool	Checksum/SHA256	OK						
	A-2	Imagefile	DFaaS tool	Checksum/SHA256	OK		X				
Conclusion: Image integrity ok with a certainty level of C5											
Time zone of	A-1	Processed data	DFaaS tool	Tool/ASCII	UTC -8		X				
computer A-1	A-1	Processed data	Computer forensic tool	Tool/ASCII	UTC -8				X		
	A-1	Registry: SYSTEM	Registry tool 1	Tool/Key value data	UTC -8		X		X		
	A-1	Processed data	Registry tool 2	Tool/ASCII	UTC 0	X	X	X	X	X	Found bug in tool
	A-1	Registry: SYSTEM	Registry tool 3	Tool/Key value data	UTC -8						
Conclusion: Time zone is UTC -8 with a certainty level of C4											
Time zone of	A-2	Processed data	DFaaS tool	Tool/ASCII	UTC +2		X				
mobile phone A-2	A-2	Processed data	Phone forensic tool	Tool/ASCII	UTC +2						
	A-2	Settings.db	Sqlite tool 1	Tool/Setting	UTC +2		X		X		
Conclusion: Time zone is UTC +2 with a certainty level of C5											
Timestamps from GoChat	A-1	Processed data	DFaaS tool	Tool/ASCII	Message sent 13:10 UTC -8		X				
	A-1	GoChat conv.db	Sqlite tool 1	Tool/Setting	Message sent 13:10 UTC -8				X		
	A-1	GoChat conv.db	Sqlite tool 2	Tool/Setting	Message sent 13:10 UTC -8		X		X		
	A-1	GoChat conv.db	Hex tool	Tool/Raw data	Message sent 13:10 UTC -8				X		
	A-2	Processed data	DFaaS tool	Tool/ASCII	Message sent 20:10 UTC 0		X				
	A-2	GoChat talk.plist	Phone forensic tool	Tool/ASCII	Message sent 20:10 UTC 0				X		
	A-2	GoChat talk.plist	Plist tool	Tool/Setting	Message sent 20:10 UTC 0		X		X		
	A-2	GoChat talk.plist	Hex tool	Tool/Raw data	Message sent 20:10 UTC 0				X		
Conclusion: Messages was sent 22:10 (UTC +2) with a certainty level of C5											
Suspect claims messages		Suspect	Interview 05.01.19	Statement	Claims it was sent earlier that day to warn the victim						
was sent 13:10		Witness	Interview 05.01.19	Statement	Claims the suspect was not at his computer earlier that day, but visiting the witness						
Conclusion: Divergent results from interviews, claim has a current certainty level of C0											