Open Universiteit

PROF. DR. IR. H.M.A. (HARM) VAN BEEK

Digital traces: clearly unclear

Van Beek HMA (2025). Digital traces: clearly unclear.
Open Universiteit, Heerlen, The Netherlands.
DOI: 10.71583/20250919hvb

Editor Dutch version: Eefje Marijt

Translated from Dutch to English with *DeepL Pro*.

The cover image was generated using *Google Gemini / Imagen 3*.

Contents

Open Universiteit

Digital traces: clearly unclear

Inaugural lecture

Delivered at the public acceptance of the position of
endowed full professor of Digital Forensics,
established by the Netherlands Forensic Institute,
at the Open Universiteit on

Friday 19 September 2025

by prof. dr. ir. H.M.A. (Harm) van Beek

Mr. Rector Magnificus, esteemed members of the Board of Trustees, dear Dean, esteemed management of the Netherlands Forensic Institute, dear colleagues, former colleagues, family, friends, and acquaintances, both here in the room and online,

Welcome here today in Heerlen, where I have the honor of delivering my inaugural lecture upon accepting the position of endowed full professor of Digital Forensics, established by the Netherlands Forensic Institute (NFI), at the Faculty of Science of the Open Universiteit. I am delighted that you have come to South Limburg, because for most of you, this is not exactly around the corner.

Introduction

As many of you know, I have been working at the Netherlands Forensic Institute in The Hague since 2009. As a computer scientist, I investigate digital traces. I help the police and public prosecutors solve criminal cases and, as a legal expert, I explain to judges how they can use digital traces as incriminating or exculpatory evidence. As a scientist, I am also involved in several research projects. This chair therefore forms a bridge between the faculties of Science and Law at the Open Universiteit, and between the Netherlands Forensic Institute and these two faculties.

I will first provide some background information. I will explain what digital traces are, how data can be converted into traces, and how these are normally used for digital investigation and evidence gathering. I will then briefly explain why traces in the digital world work differently than in the physical world and what this means for the reliability of investigations. After all, it is important that evidence is interpreted correctly. Next, I will discuss the scientific research that I want to develop within the chair in order to better interpret digital traces and use them lawfully.

1 Digital Traces

Nowadays, we do a lot digitally. We communicate digitally, take photos digitally, and find partners digitally. Or we receive digital assistance, for example, in traffic when finding the right route, or in stores when making contactless payments. Consciously or unconsciously, we are constantly leaving digital traces behind. And not just a few, no, we scatter them around as if it costs nothing. To get a good idea of this, I would like to conduct a little experiment with you. If you have a cell phone with you, would you raise your hand?

As you can see, this applies to virtually everyone, regardless of age, gender, job, or role. All of these phones contain traces that are related to this inaugural lecture. Consider for yourself how many of the following questions you would answer “yes” to:

- Did you receive an email or message about this lecture?
- Is this lecture in your digital calendar?
- Did you search the internet for more information about inaugural lectures?
- Or did you discuss it with chatGPT?
- Did you send or receive a message about planning your visit?
- Did you use a navigation system or app to get to Heerlen?
- Did you make a contactless payment with your phone on the way here?
- Did you take and/or share a photo just now or on your way here?
- And do you perhaps wear a smartwatch or fitness tracker?

With every “yes”, you leave digital traces, namely the messages and photos themselves, the destinations in the navigation app, and transactions in the payment app, but also the data that is automatically collected, such as your location, the number of steps you have taken, and how full your phone’s battery is. So, consciously and unconsciously, we all leave digital traces behind, and together we have left many thousands of traces that can be traced back to this speech.

These traces can be roughly divided into three different groups. Firstly, there are traces that we as humans consciously create, in which we record and/or share information with each other. Examples include chat messages, emails, and digital photos. Secondly, there are traces that are recorded by sensors in devices without user intervention, such as the location of your phone, how many steps you have taken, or your heart rate. And thirdly, there are traces that record usage and settings, such as traces of unlocking your phone or connecting to a Wi-Fi network. Combinations are also possible, such as a digital photo that also records a location.

All these traces can say something about the use of the device and therefore also about the user of the device. If the user is potentially involved in a criminal offense, for example as a suspect or victim, these traces can also provide clues. They can thus help to solve a case.

2 Digital Investigation in Practice

Suppose that in two weeks' time, the police suddenly show up at your doorstep. You are suspected of stealing some products. This took place on September 19, 2025, around 4:15 p.m., from the supermarket where you usually do your shopping.

Traces on your phone could potentially confirm or refute this and thus help the police solve the case. Perhaps you sent a message about it or even recorded the theft yourself. Sensor traces such as the location of your phone could also provide clues. You will probably deny it, because at that time you were at this lecture at the Open Universiteit in Heerlen. And, depending on how often you just answered "yes," there are traces of this in your phone, car, or smartwatch.

If the police want to use data from your phone, this is done through a fairly fixed process, which must take into account technical, forensic, and legal requirements [6].

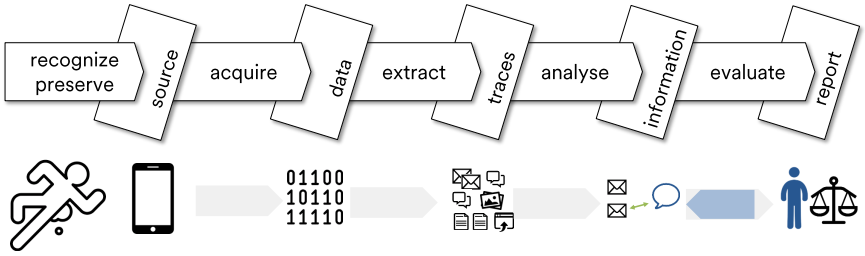

Figure 2.1: Steps in a forensic investigation [6]

2.1 Securing Digital Data

If your phone is confiscated, special equipment will first be used to extract all the *data* from it. This data consists of all the zeros and ones stored on hard drives and memory chips. This is unlikely to happen for shoplifting, but there may be more to it than that. If there is sufficient reason for a search, the police will quickly find dozens of devices containing potential evidence in addition to your phone.

I don't know about you, but I have a USB stick or old phone in almost every drawer, and I'm sure you have a computer or laptop too.

After seizure, the devices are given to digital experts to extract the data. This is quite an art in itself, especially with all the security measures that ensure that the data in those devices is not easily accessible to unauthorized persons. All these devices therefore require their own techniques to extract the data. If successful, this yields a huge amount of data per device.

2.2 Making the Data Insightful and Searchable

Once the data has been collected, the next step is to make it insightful and searchable, i.e. to turn it into traces that we as humans can understand. There are various ways of doing this, depending on where the data comes from, how much data there is, and the research questions of the investigators working on the case. In general, the data is opened by digital experts in forensic computer programs, often focused on a specific source such as a phone or a USB stick or hard drive from a computer.

NFI builds powerful forensic search engine for digital investigations [17]
The average number of digital traces that investigators find in criminal cases doubles every fifteen months. To keep these growing amounts of data quickly and easily searchable, the Netherlands Forensic Institute (NFI) developed the forensic search engine Hansken.

With the help of Hansken, the police can continue to carry out digital forensic investigations quickly and efficiently. It does not matter whether investigators have to search through a number of laptops or an entire server park. All seized data is copied into Hansken. The software then identifies as many traces as possible. Thanks to its structure, Hansken, like a search engine, is able to recognize and make searchable an enormous amount of data.

This Dutch implementation of *Digital Forensics as a Service* [2, 3, 7] is the successor to Xiraf [1]. Partners from the international community of intelligence and investigation services are now also actively contributing to the platform. The academic network surrounding Hansken is also growing. Participating universities and colleges are collaborating by using Hansken for educational and research purposes.

In 2024 Hansken won first prize in the category 'Digital Transformation' at the European Public Sector Awards [21].

See hansken.org for more information about Hansken.

Digital experts can use these computer programs to convert data into readable traces, such as emails and images, and then search through them. Digital experts often have difficulty determining which data is relevant because they are not involved in the case. Case investigators, who are unfamiliar with the content of the seized devices, therefore often ask for all available emails, photos, or chat messages, and sometimes even just “everything.” This is neither quick nor efficient. One solution is for investigators themselves, with the help of digital experts, to use forensic computer programs to search for interesting traces.

If there are many different sources, the investigators working on a case want to search all the data from all sources at the same time. They are looking for clues or evidence, but it doesn’t matter to them whether that one email was sent using Outlook from a computer, the Mail app from an iPhone, or Gmail from a browser. To make this possible, I developed a platform together with colleagues from the NFI and several intelligence and investigation services.

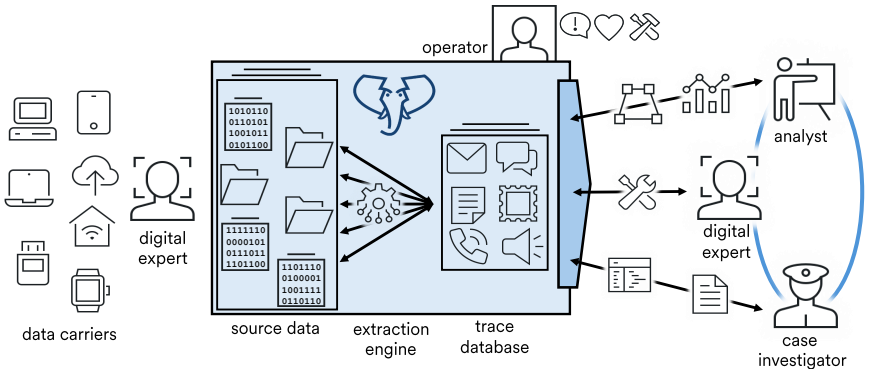

Figure 2.2: Hansken: Digital Forensics as a Service [2, 3, 7]

This platform, called Hansken [17], uses techniques that together ensure that all traces in the data are searchable. And that this search is in line with how investigators are used to working. Searching on the Hansken platform is similar to searching for a product in an online store. You open the Hansken website, log in, open the case you are working on, and you will see all the traces. As explained earlier, there are a lot of them. In an average case, you will find as many as 10 million traces.

Of course, not all of them are relevant to the investigation, just as not all items of clothing in a web store are of interest to you. There, you can filter for T-shirts, for example, made of a certain material, in a certain color, in a certain size, with a certain print. At some point, you start assessing them one by one. In a similar way, investigators can use Hansken filters to ultimately view individual traces. For example, they can filter on chat messages, emails, or images. They can also combine filters, such as all chat messages sent in 2025 that contain the word "inaugural."

Nowadays, investigators are also assisted in this by artificial intelligence (AI) [12]. Hansken can, for example, determine what is in an image. Hansken can also convert spoken messages into readable text. Investigators can then search for photos of a firearm, a drug lab, or cash, or they can search for audio messages in which a certain name is mentioned. Modern language models can also be used to summarize a long chat conversation [9, 15], for example. This makes it easier to find potentially relevant clues.

Replica skull Hansken from the Rembrandt House to NFI [19]
She is a famous elephant, captured on canvas by the artist Rembrandt in the seventeenth century, and she embodies the future of crime fighting in the 21st century: Hansken. A replica of Hansken's skull was transferred by the Rembrandt House Museum in Amsterdam to the Netherlands Forensic Institute (NFI) in The Hague. The NFI named a digital search engine for quickly searching large amounts of digital data after the famous elephant. Investigation services from all over the world now use this search engine.

The NFI is delighted with the replica, says Harm van Beek, senior digital researcher at the NFI. He is one of the founding fathers of the Hansken digital forensic data analysis platform. "I had already seen the impressive skeleton in Florence. This is a replica of the skull, but it still gives an impression of how impressive the animal was." Van Beek was one of the people who named the platform. "The predecessor of Hansken was Xiraf. That was an acronym, but phonetically it sounds like giraffe. We wanted a similar name for Xiraf's larger successor," he says. "An elephant is a good-natured hoofed animal, just like a giraffe. We also wanted a link with the Netherlands, and a colleague came up with the name Hansken. Hansken was trained to perform tricks and, according to tradition, she could point out crooks in the audience with her trunk. She was also trained to handle weapons. We all thought that was appropriate," he laughs.

2.3 Reporting Relevant Results

The relevant traces are reported in an official report. If Hansken has been used, a report of the trace is often also extracted from Hansken. This report contains both the content of the trace, for example a chat conversation, and the location where it was found on a device. This is important for its use as evidence. In addition, the techniques used to arrive at these results are described. This allows you to check whether the data processing was carried out correctly. Even then, there may still be discussion about the reported traces. I will come back to that later.

Ultimately, this results in a file containing all the information that supports what the police believe to have happened. The public prosecutor uses this file to substantiate his view of the truth and to propose an appropriate sentence to a judge. You, as the suspect, may have a different view. Your lawyer may be able to provide exculpatory evidence, such as a witness who supports your version of events. If your exculpatory evidence is on your phone, this can be difficult. Your phone may still be with the police. The police can search for your evidence and record it in an official report. In some cases, your lawyer can now also access the data from your phone [20] and ask the public prosecutor to include specific evidence in the case file. So that chat message saying you were "just in time for the lecture in Heerlen" could prevent you from getting into trouble.

NFI enables access to crypto communications in lawyers' workplaces [20]
The Netherlands Forensic Institute (NFI) has developed a method in collaboration with the police and the Public Prosecution Service (OM) that allows lawyers to view encrypted communications, such as those using Encrochat, from their own offices using the Hansken digital search engine. "This new method of accessing digital evidence using a professional search engine from lawyers' own workstations is unique in the world," says Martijn Egberts, national public prosecutor for digital investigation. Initial feedback from the legal profession has been positive.

With the new method, lawyers can access the data in Hansken from their own computers via a secure connection, after obtaining permission to view the data. Lawyer Desiree de Jonge tried out the new method: "The implementation of Hansken's remote access function is a major and important step in actually facilitating the access options that the defense must have in criminal cases."

3 The Complex Digital World

The above approach and investigation of digital traces seems obvious. What could possibly go wrong? The answer to that question is: a lot!

This has everything to do with the differences between the digital world and the physical world, and therefore traces in those worlds [16]. The digital world was conceived and designed by humans. That is why, for example, the laws of nature do not always apply and this world is changing very rapidly.

3.1 A Variety of Traces

For many things we used to do in the physical world, we now have a digital alternative, which has sometimes even made the old ways obsolete. This continuously creates new types of digital traces.

We used to write letters, make calls with a landline phone, or send a fax. Now we communicate via WhatsApp, email, and social media. We used to use film rolls, have them printed, and store the photos in photo albums; now we take digital photos with our smartphone cameras and store them “in the cloud.” We used to read books from the library or bookstore; now we also read e-books on tablets or e-readers and listen to audiobooks. We now do our banking via an app; in the past, you had to go to the bank or fill out a giro form. We now hold meetings from home via video link. We shop in online stores and are going to high street shops less and less. In the past, we had cassette tapes, CDs, and vinyl records; nowadays, we listen to music via streaming services, just like movies and series. We used to rent videotapes from the video store. And so on. All these new ways of doing things create new traces that we don’t all understand and can interpret.

Not only is everything digital now, but the digital world is also constantly changing. Every new device or new or updated app can lead to different or new traces. Techniques that we can use today to gain insight into data may not work tomorrow.

NFI designs international knowledge base on digital forensic investigation [23]
The Netherlands Forensic Institute (NFI) has collaborated with the universities of Oxford and Lausanne to develop a framework for an international knowledge base for digital forensic investigation. Harm van Beek of the NFI contributed to the design of the knowledge base: "Currently, there is no central location for gathering knowledge about how to conduct digital forensic investigations. The knowledge base will record the various steps in the forensic digital process in a clear and structured manner. The knowledge base will give digital forensic investigators more control over knowledge and developments in the constantly evolving world of digital forensic techniques."

The knowledge base is called SOLVE-IT [14, 23], which stands for Systematic Objective-based Listing of Various Established (Digital) Investigation Techniques. In addition to the steps in the process, it also records which techniques you can use, what the vulnerabilities are, what you can do to limit them, and refers to sources that provide more information.

See github.com/SOLVE-IT-DF/solve-it for more information.

3.2 Different Naming of Traces

A second challenge is that we use different terms for the same thing. For example, one app calls an image a 'picture', another calls it an 'image', and yet another calls it a 'photo'. If you are an investigator looking for all images, you want to find traces from all three apps. This calls for a standard: we need to agree on how we name digital objects.

Many standards already apply in the digital world, such as standards that define how you can send data packets over the internet. We also agreed on an initial standard for emails a long time ago, in 1982 to be precise [25]. When you send an email, it doesn't matter whether the recipient uses Outlook, Gmail, or iCloud Mail. These apps all follow the email standard. This is not (yet) the case for chat. For example, you cannot currently send a message from WhatsApp to Signal.

Despite these standards, one app still calls an email a 'mail', while another app calls it a 'message'. And one app calls the sender the 'from', while another app calls it the 'sender'. There are several initiatives to achieve uniform naming, at least for digital forensic investigation. Hansken has a uniform trace model for this. A larger international initiative is CASE, in which I am involved, which is partly based on the concepts of Hansken's trace model.

NFI one of the founders of universal cyber language that facilitates international fight against crime [18]
To effectively combat crime, international cooperation between investigative services and digital forensic investigators is essential. Consider, for example, the exchange of forensic tools for reading encrypted phones. To facilitate this cooperation, it is useful for the tools to record all traces in the same 'cyber language'. An international language has now been developed for this purpose: Cyberinvestigation Analysis Standard Expression (CASE) [4, 5, 18]. The Netherlands Forensic Institute (NFI) is one of the founders of this global language.

See caseontology.org for more information about CASE.

3.3 Multiple Originals of a Single Object

Unlike physical objects, objects in the digital world can be in multiple places at the same time. For example, a chat message you have sent is stored on your phone and simultaneously on the recipient's phone. Or the email with the invitation to this lecture is stored on all your computers and phones at the same time. In addition, data is stored "in the cloud" in different, often unknown locations at the same time. Do you know where your email is stored? Is it in the Netherlands, abroad, or both?

If there are multiple suspects in an investigation, we also see traces appearing multiple times in a case. This does not make it any easier for investigators, because after filtering, they sometimes find the same message five times, or the same image ten times. And they are often not entirely identical in detail. For example, because a chat message was sent from an iPhone to an Android phone, or because a photo was cropped via WhatsApp. The traces then appear similar, but are recorded slightly differently in the data. An open question is therefore how we can generally identify digital traces as unique, and how we can determine whether traces are the same digital object.

3.4 Multiple Perspectives on Traces

When you apply filters, you are looking for all digital traces that match your filter. But digital traces often take different forms, depending on how you look at them. For example, an image can also appear as an email attachment. So if you filter for images, you want to find them. But if you filter for email attachments, you want to find them too. This phenomenon has led to Hansken's trace model, as well as CASE, describing traces from multiple perspectives simultaneously.

3.5 National Borders Do Not Exist

Another problem in the digital world is that national borders do not exist. Formally, they do exist, because a device is subject to the laws of the country in which it is located. The same applies to the data stored on it. For data stored "in the cloud," complex algorithms determine where this data can best be stored. In doing so, the algorithms, consciously or unconsciously, do not always take national borders into account. This is a major problem for investigative services. Within our own borders, we are responsible for our own legislation. But digitally, you can easily cross national borders without supervision. For example, a hacker from the Far East can steal digital coins in the Netherlands via digital means [11]. Bringing this hacker to justice in the Netherlands is often an impossible task. Cybercrime, but certainly also organized serious crime, is an international phenomenon, made possible in part by all the digital facilities available.

4 Reliability of Investigative Results

Standardization and filtering work well with traces that we as humans have created, such as chat messages and photos. This provides tactical information that, in light of the case, either supports or refutes a particular view. As mentioned, there can be a lot of debate about digital traces. This debate may concern the process of processing the data, the end result of this processing, or the significance of these results in the context of the case.

4.1 Understanding What the Data Means

The data must be presented in such a way that we can conduct research on it or that analysis software can use the data for (follow-up) analyses. Files such as images and documents are displayed on the screen by computers and phones in a format that is understandable to humans. However, on the data carriers themselves, these are digitally encoded and structured, or "translated into strings of zeros and ones."

The encoding of the data—the recording of data in a series of zeros and ones—and the structure of the data—how these series are combined—determine what the data means. This "translation" takes place through the programs and apps used to record the data, such as the camera app, the Office program, the email program, or the chat app.

Because software is constantly evolving—programs and apps are updated very regularly—the behavior of the software also changes. Usually, this has no impact on how data is encoded and structured, but existing and new data is regularly recorded in new or different endocings and/or structures as a result.

Some of the encodings and structures used are known, but others are not. This mainly depends on which program or app was used to record the data.

Programmers write software in program code. This program code is called the source code. In order to execute this code on a computer, it is converted (compiled) into computer instructions. This executable code is no longer readable by humans. If the source code of the program or app is available (open-source software), its operation is transparent and the coding and structure can be traced.

It also happens that the data is encoded and structured according to agreed standards. And some encodings and structures are used more than others. They then become a de facto standard, or the first choice for developers of new or modified software. Examples are ASCII for text encoding, JPEG for images, JSON for data, and SQLite for databases. For well-known encodings and structures, there are often methods to make them transparent. And the effectiveness of these methods is often well known. If the source code of the software is not available and there is no documentation, the encodings and structures are unknown. There are roughly four ways to make this data transparent.

Firstly, the software that recorded the data can be used. This provides insight in the data, even though the encoding and structure of the data are still unknown. Investigators can then access the data in the same way as a normal user working with the software. This can be done, for example, with complex databases from bookkeeping programs. The bookkeeping can then be opened and examined using the bookkeeping program.

A second possibility is to investigate the behavior of the software through experiments. By comparing known input data with the data recorded during the experiments, it is possible to deduce how the data is encoded and structured. For example, if you send a chat message to a known phone number, you can identify where the message and phone number appear in the data.

Thirdly, we can examine the internal workings of the software. By following the computer instructions of the executable code step by step, we can find out how the software behaves. This is called reverse engineering. AI is also increasingly being used to analyze code these days.

The fourth option is to apply all kinds of known decoding methods. By trying out various known decoding methods on the data, it may be possible to make the data readable. This is called brute-forcing. AI also helps here, for example by recognizing patterns in the data.

Forensic investigation programs can then be developed to provide insight into the data based on the knowledge gained. The results of these programs can ultimately serve as evidence in court. Because it is often not possible to investigate all the possibilities of the software and assumptions may have to be made, such a program may produce incomplete or incorrect results. It is therefore important that the functioning of these programs is known [8]. If this is not the case, we refer to such a program as a 'black box'. Checking with other tools, including the software that encoded and structured the data, is therefore almost always desirable. The results of such a program are then compared with the results of another program. This is called *dual tool verification* [26, 29] and is a commonly used method for checking results from forensic investigation programs. The idea behind this is that if different investigators have independently developed their own programs to analyze data and these different programs produce the same result, this result is verified.

4.2 What the Evidence Reveals in a Criminal Case

In criminal cases, everything revolves around truth and reliability. This involves answering questions such as who, what, when, how, and why.

Evidence is not an end in itself, but a tool for reaching a conclusion: did the suspect do it? And if so, how? In law, we use a nice word for this: *conclusiveness*. By this we mean: to what extent does a particular piece of evidence support the conclusion we want to draw?

This is not always as black and white as it seems. Was the chat message sent by the suspect or could someone else have done it? Was the email sent from the phone or another computer? When was the website visited, before or after the crime? Where was the photo taken and with what device? In short: there is a trace, but how conclusive is it?

The above questions can be asked about virtually all digital evidence, and you can probably think of many more yourself. To answer these questions, forensic experts not only examine the traces found, but often also the more technical information recorded around the traces in devices. This includes settings and details of the historical use of the devices and apps. This is research based on what science teaches us. The motto of the NFI is therefore 'focused on truth, guided by science, for a safer society.'

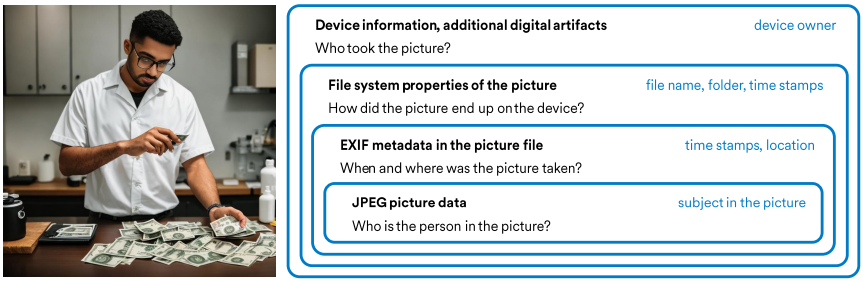

Figure 4.1: Criminalistic zoom levels for a picture on a phone [16]

Depending on the question, scientific literature is consulted, experiments are conducted to reconstruct similar traces, and the traces are further investigated using available or newly developed forensic programs.

These questions can be asked for various reasons. The answers can provide direction during the investigation, for example in finding possible suspects or establishing possible scenarios. We refer to this as steering information. Traces can also serve as evidence in court. In that case, as forensic experts, we make a statement about the traces in light of various hypotheses. Your chat message about ‘being just in time for this inaugural lecture’ carries a certain weight in the scales of Lady Justice, when weighing up whether or not you stole something from the supermarket. That is also quite difficult, because ultimately the message must also be linked to you personally. Were you the one who sent that message with your phone? Or was it someone else?

5 The Chair in 'Digital Forensics'

It has become clear to you that there are many snags involved in correctly interpreting digital traces. This applies to obtaining data from devices, translating that data into traces that we as humans can understand, and interpreting traces in the context of a case. Through research within the chair, I want to tackle a number of these topics, but certainly not all of them. The focus of the research is on interpreting digital traces within the legal framework and context of a criminal case. This chair is therefore at the intersection of computer science, information science, and law.

The collaboration between the NFI and the Open Universiteit is essential in this regard, particularly with and between the departments of Computer Science, Information Science, and Public Law (Criminal Law section).

5.1 Interpreting Using Formal Methods

I have a background in modeling processes using formal methods. These techniques are still rarely used for (digital) forensic investigationz, even though they contain many elements that are particularly important for detection and evidence gathering in court. But what are formal methods? Formal methods are a kind of super-accurate way of reasoning. They are techniques from computer science and mathematics that allow us to describe, check, and/or analyze systems—such as software, networks, or (digital) processes—down to the smallest detail. You can compare it to the difference between a rough sketch and a construction drawing: the former gives a general idea, while the latter explains how everything fits together.

Formal methods use mathematical logic and models to analyze computer systems or processes. They have been used for years in aviation and railways, for example. Think of an autopilot or a system that controls trains. In these cases, you want to be absolutely sure that everything is working properly, because even a small error can have major consequences. That is why these systems are often formally analyzed: Does it do what you expect in all conceivable circumstances? Are there situations in which

the system could crash or, worse still, cause an accident? A well-known example comes from space travel. NASA uses formal methods to prevent errors when developing software for spacecraft. For example, for the system that controls the Mars rover Curiosity [28]. You can't quickly send an update to it if something goes wrong, so you have to be sure that everything is correct beforehand.

But what if something did go wrong? Then we want to understand afterwards how it went wrong. Formal methods are also used for this. And that's where digital forensic investigation comes in. In digital forensic investigation, we try to find out what happened: Who did what, when, and how? Like the shoplifting you are suspected of. The traces you leave behind are complex and not always clear-cut. This is where formal methods come in handy. They help us to better understand, structure, and analyze those digital traces. They force us to reason precisely and transparently: What do we know for sure based on the traces? What could have happened? And what can be ruled out? Formal methods help with this by making assumptions explicit and therefore verifiable.

5.1.1 Reconstructing Timelines

One example where this use of formal methods can be useful is in creating timelines. Imagine: the police find a number of chat messages, photos, and log entries on your phone. They are time-stamped, but some of the timestamps are unreliable (for example, because the system clock was set incorrectly) [13]. By using formal models of time—for example, interval logic—we can construct a consistent timeline that takes the uncertainties into account. We can then say: "If this photo was taken after this chat message, then this action must have taken place between approximately 4:00 p.m. and 4:05 p.m." These time analyses are essential for comparing scenarios: is your alibi correct? Does your story match what we can reconstruct digitally?

5.1.2 Modeling System Behavior

Another example is the behavior of an app or a network. Suppose the police suspect that your chat was not sent by you but automatically by the app. As a researcher, you want to be able to determine whether this is technically possible. By creating a formal model of how that app works—which actions lead to which traces—we can verify whether that

scenario is plausible. Is there evidence of automatic behavior, or were human actions required? These arguments are strong because they are based on logic rather than on unrecorded assumptions.

5.1.3 Testing Scenarios

Formal methods can also help to test the available digital evidence against different scenarios. For example, we can automatically check which traces match the scenarios, i.e., which traces are relevant. Suppose there is a debate about whether a USB stick was connected to the suspect's laptop. We can then formally model what traces a USB connection leaves behind and when. This allows us to rule out certain scenarios or, conversely, make them plausible.

Formal methods are not a magic formula, but rather a smart way to carefully piece together digital puzzles. In this way, they help to remove ambiguities from the discussion. They make it possible to ask questions such as: What do we know for sure? What don't we know? Which facts fit the scenarios? These types of questions help not only investigators, but also police, prosecutors, judges, and lawyers to make better choices and decisions. They make the analysis repeatable, transparent, and verifiable. This is not only good for finding the truth, but also for the legal certainty of suspects.

5.2 Lawful Use of Digital Traces

Your devices may contain valuable clues for investigative services. They may contain crucial evidence in a criminal case. But that does not mean that everything can simply be used for an investigation. In law, we call this the 'legality test'. This involves a few key concepts. First, *proportionality*: the remedy must not be worse than the disease. Making a complete copy of all the data on your phone is probably not proportional if you are suspected of a single shoplifting offense. Second, *privacy*: Your personal privacy must be respected, also in a criminal case [24]. For example, is it necessary to investigate everyone in your address book? Or to view all your private photos from the past few years. And thirdly, the *legal professional privilege* of lawyers, notaries, doctors, or clergy, with whom you share information in confidence. 'In confidence' means that the police are not allowed to view this communication either.

The reality is that there is often a huge amount of data, which Hansken, for example, has made understandable and searchable. Investigative services have to search through these enormous amounts of data to find that one piece of evidence. And that is where the risk lies: if you search everything "just in case," you will quickly find yourself doing disproportionate work. What's more, if you also view all communications with a doctor or lawyer, you run the risk that the evidence you do find may not be used later because it was obtained unlawfully. This is not only inefficient, but also bad for confidence in the rule of law.

Formal methods can play a role in searching large amounts of digital data in a smart, precise, and verifiable way, without seeing more than is necessary.

5.2.1 Responsible Filtering

Sometimes digital data contains communications with persons who are subject to legal privilege, such as lawyers. This data may not be accessed, and certainly not used in the investigation [10, 22]. Both formal methods and AI can help to automatically recognize such communications, based on metadata (such as the lawyer's email address), communication patterns, or even stylistic characteristics. Forensic programs can automatically mark or block this data before it is seen by a human. In this way, you can not only respect the right to legal professional privilege, but also enforce it technically. However, this requires further research, as it must be done in a transparent and responsible manner.

5.2.2 Dealing with Data Proportionally

In the case of large-scale digital seizures—for example, the seizure of servers or cloud storage—it is important not to do more with the data than is necessary. But this also applies to smaller data collections.

In 2017, the Dutch Supreme Court concluded in the smartphone ruling [27] that the police may only view a smartphone to a limited extent. A full search of private data requires the permission of a public prosecutor or judge. Following a ruling by the European Court of Justice in the case CG/Landeck [30], the Dutch Supreme Court ruled in March of this year that when investigating electronic data carriers, "no greater infringement of the user's privacy than is necessary may be made" [31].

To ensure this, the investigation must be carried out automatically as far as possible.

Formal models can be used to evaluate search strategies in advance: which data is relevant, and which is not? Is it really necessary to go through the entire archive? Can you build in an intermediate step, for example by first looking only at metadata? By formally modeling these questions in advance, you can organize the search process in an efficient and targeted manner, ensuring that it remains proportionate.

5.3 A Strong Criminal Justice System

The beauty of this formal approach is that it also helps the criminal proceedings themselves. Evidence that has been obtained in a careful and lawful manner is much stronger. It is more likely to stand up in court. And it increases the confidence of citizens, suspects, and victims in a fair trial.

Formal methods are therefore not an obstacle to investigation, but an opportunity: they offer ways to make the work more precise, fairer, and more transparent. And that is precisely what the digital rule of law is all about. This research program is ambitious and necessary given the central role that digital traces play in modern crime fighting. By combining the forces of computer science, information science, and legal studies and connecting them with the practice of the NFI, this chair can strengthen the search for truth and the administration of justice in the Netherlands and beyond.

I look forward to implementing this research program together with colleagues from the Open Universiteit, the NFI, and other partners, thereby making a scientific contribution to a safer society.

6 Acknowledgements

Finally, I would like to thank a number of people.

First of all, I would like to thank the Executive Board of the Open Universiteit, and in particular Rector Magnificus Professor Theo Bastiaens, for the trust they have placed in me and for appointing me as endowed professor of Digital Forensics. The same applies to the management of the NFI, general director Marc Elsersohn and director of Science & Technology Annemieke de Vries.

I would never have been able to hold this chair if Professor Jos Baeten and Professor Sjouke Mauw had not convinced me in 2000 to continue my graduation work in a PhD research project. Thank you for the confidence you had in me at the time. And Jos, of course, extra thanks for your contribution this afternoon at the symposium.

I started that research at a time when I had just launched an internet startup with a number of friends. Max Hufkens and Mark Hoogendoorn, as well as all my other former colleagues at ISAAC, thank you very much for giving me the space I needed during that challenging period. And, of course, for understanding that after ten years of working together, I wanted to take my career in a new direction toward (applied) science.

I would also like to thank all my colleagues and former colleagues at the NFI, who have always been so helpful. Especially my colleagues at Hansken, in the Digital & Biometric Traces division, but also those in the various support services. And the people I work with in the various organizations in the criminal justice chain. There are too many to mention individually here. I would like to highlight a few, however. Ruud van Baar and Erwin van Eijk, because you dared to take the plunge with me and took on the responsibility of developing Hansken together. Hans Henseler, with you I have been able to give more shape to my scientific ambitions in recent years. Many thanks for that. Jeroen Venema, because with your positive energy you made this afternoon's symposium a success. And of course Martijn Egberts, for your support for all the work we do at the NFI and of course also for your contribution earlier today.

Of course, I would also like to thank my new colleagues at the Open Universiteit. Vincent van der Meer, thank you for your help in the preliminary stages of my appointment and for putting me in touch with the dean of the Faculty of Science, Professor Petra de Weerd-Nederhof. Petra, from the moment we first met, you placed your trust in me and supported me in the process of obtaining this special chair at the Open Universiteit. Your *can-do* attitude is infectious! Hugo Jonker, you are the person I have the most contact with here at the Open Universiteit. I enjoy our discussions and appreciate your support in filling the chair. Professor Wilma Dreissen, thank you for your participation in my board of trustees and your presentation this afternoon. And, of course, many thanks to my colleagues who gave lightning talks this afternoon about the various research projects within the Computer Science department.

Many thanks also to my family and in-laws. My parents taught me that hard work pays off. And as the tile in your bathroom in Westerhoven says: *"You should think in terms of solutions, not problems."* Together with the rest of my family, you ensure that I keep both feet firmly on the ground. In addition, you have always given me the space to follow my own path! Without that space, I would not be here today.

But my greatest thanks, of course, go to my family, to Cindy and my children. You always support me! Cin, somehow you manage to keep the family running smoothly and give me the space I need, so that I can travel here to Heerlen regularly, for example. We teach our children that above all they should enjoy themselves and have fun. So I appreciate all your jokes, even about this ceremonial dress code. Thank you so much for your support and love.

I said.

References

Scientific Publications

[1] R. Bhoedjang, A. van Ballegooij, H. van Beek, J. van Schie, F. Dillema, R. van Baar, F. Ouwendijk, and M. Streppel, "Engineering an online computer forensic service," *Digital Investigation*, vol. 9, no. 2, pp. 96–108, 2012. DOI: 10.1016/j.diin.2012.10.001.

[2] R. van Baar, H. van Beek, and E. van Eijk, "Digital forensics as a service: A game changer," *Digital Investigation*, vol. 11, S54–S62, 2014, Proceedings of the First Annual DFRWS Europe. DOI: 10.1016/j.diin.2014.03.007.

[3] H. van Beek, E. van Eijk, R. van Baar, M. Ugen, J. Bodde, and A. Siemelink, "Digital forensics as a service: Game on," *Digital Investigation*, vol. 15, pp. 20–38, 2015, Special Issue: Big Data and Intelligent Data Analysis. DOI: 10.1016/j.diin.2015.07.004.

[4] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, and A. Nelson, "Advancing coordinated cyber-investigations and tool interoperability using a community developed specification language," *Digital Investigation*, vol. 22, pp. 14–45, 2017. DOI: 10.1016/j.diin.2017.08.002.

[5] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, and A. Nelson, "The evolution of expressing and exchanging cyber-investigation information in a standardized form," in *Handling and Exchanging Electronic Evidence Across Europe*, M. A. Biasiotti, J. P. Mifsud Bonnici, J. Cannataci, and F. Turchi, Eds. Cham: Springer International Publishing, 2018, pp. 43–58. DOI: 10.1007/978-3-319-74872-6_4.

[6] H. van Beek, "A forensic visual aid: Traces versus knowledge," *Science & Justice*, vol. 58, no. 6, pp. 425–432, 2018. DOI: 10.1016/j.scijus.2018.08.006.

[7] H. van Beek, J. van den Bos, A. Boztas, E. van Eijk, R. Schramp, and M. Ugen, "Digital forensics as a service: Stepping up the game," *Forensic Science International: Digital Investigation*, vol. 35, p. 301021, 2020. DOI: 10.1016/j.fsidi.2020.301021.

[8] H. van Beek, "Forensic safeguards in Hansken (NL)," Netherlands Forensic Institute, Information sheet, Nov. 2021.

[9] H. Henseler and H. van Beek, "Chatgpt as a copilot for investigating digital evidence," in *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*, 2023, pp. 58–69.

[10] H. van Beek, “Destruction of digital traces containing privileged information (NL),” Netherlands Forensic Institute, Information sheet, Apr. 2023.

[11] R. Argentini and H. van Beek, “Bitcoin (NL),” Netherlands Forensic Institute, Information sheet, Jan. 2024.

[12] H. van Beek and H. Henseler, “Servicing digital investigations with artificial intelligence,” in *Artificial Intelligence (AI) in Forensic Sciences*, Z. Geradts and K. Franke, Eds., first edition, John Wiley & Sons Ltd, 2024, pp. 103–122.

[13] C. Vanini, C. Hargreaves, H. van Beek, and F. Breitinger, “Was the clock correct? exploring timestamp interpretation through time anchors for digital forensic event reconstruction,” *Forensic Science International: Digital Investigation*, vol. 49, 2024, DFRWS USA 2024 - Selected Papers from the 24th Annual Digital Forensics Research Conference USA. DOI: 10.1016/j.fsidi.2024.301759.

[14] C. Hargreaves, H. van Beek, and E. Casey, “Solve-it: A proposed digital forensic knowledge base inspired by mitre att&ck,” *Forensic Science International: Digital Investigation*, vol. 52, Mar. 2025, DFRWS EU 2025 - Selected Papers from the 12th Annual Digital Forensics Research Conference Europe. DOI: 10.1016/j.fsidi.2025.301864.

[15] G. Michelet, H. Henseler, H. van Beek, M. Scanlon, and F. Breitinger, “Fine-tuning large language models for digital forensics investigations: Case study and general recommendations,” in *Proceedings of the 14th International Conference on IT Security Incident Management & IT Forensics (IMF 2025)*, Sep. 2025. DOI: 10.1145/3748264.

[16] H. van Beek, C. van den Pol, and J. van der Weerd, “Criminalistic zoom levels: Unravelling the hierarchy of forensic traces,” *Forensic Science International*, vol. 372, Jul. 2025. DOI: 10.1016/j.forsciint.2025.112498.

Media

[17] Netherlands Forensic Institute. “NFI developed forensic search engine for digital investigation.” (Oct. 2015), [Online]. Available: https://www.forensicinstitute.nl/news/news/2015/10/14/nfi-developed-forensic-search-engine-for-digital-investigation.

[18] Netherlands Forensic Institute. “NFI one of the founders of universal cyber language that facilitates the international fight against crime (NL).” (Feb. 2022), [Online]. Available: https://www.forensischinstituut.nl/actueel/nieuws/2022/02/02/nfi-een-van-de-grondleggers-van-universele-cybertaal-die-internationale-strijd-tegen-criminaliteit-makkelijker-maakt.

[19] Netherlands Forensic Institute. “Replica skull hansken from rembrandt house to NFI (NL).” (Feb. 2022), [Online]. Available: https://www.forensischinstituut.nl/actueel/nieuws/2022/02/021/replica-schedel-hansken-van-rembrandthuis-naar-nfi.

[20] Netherlands Forensic Institute. “NFI enables access to crypto communications in lawyers’ workplaces (NL).” (Mar. 2023), [Online]. Available: https://www.forensischinstituut.nl/actueel/nieuws/2023/03/20/nfi-maakt-inzage-cryptocommunicatie-op-werkplek-advocaten-mogelijk.

[21] Netherlands Forensic Institute. “Digital forensic platform hansken wins european public sector award.” (Mar. 2024), [Online]. Available: https://www.forensicinstitute.nl/news/news/2024/03/21/digital-forensic-platform-hansken-wins-european-public-sector-award.

[22] Netherlands Forensic Institute. “Digitaal platform Hansken helpt bij filteren van toegang tot informatie op digitale gegevensdragers.” (May 2025), [Online]. Available: https://www.forensischinstituut.nl/actueel/achtergrondverhalen-nfi/digitaal-platform-hansken-helpt-bij-filteren-van-toegang-tot-informatie-op-digitale-gegevensdragers-eafs-2025-deel-3.

[23] Netherlands Forensic Institute. “NFI co-creates international digital forensics knowledge base.” (May 2025), [Online]. Available: https://www.forensicinstitute.nl/news/news/2025/05/02/nfi-co-creates-international-digital-forensics-knowledge-base.

[24] Open Magazine, Open Universiteit. “Privacy and digital investigation, a difficult issue (NL).” (May 2025), [Online]. Available: https://www.ou.nl/-/openmagazine-2025-vranken-dreissen-van-beek.

Other References

[25] D. Crocker, *Standard for the format of ARPA internet text messages*, RFC 822, Aug. 1982. DOI: 10.17487/RFC0822.

[26] Association of Chief Police Officers (ACPO), "Good Practice and Advice Guide for Managers of e-Crime Investigation," ACPO Managers Guide, Sep. 2010.

[27] Supreme Court of the Netherlands, *Smartphone ruling*, ECLI:NL:HR:2017:584, Apr. 2017.

[28] R. Cardoso, M. Farrell, M. Luckcuck, A. Ferrando, and M. Fisher, "Heterogeneous verification of an autonomous curiosity rover," in *NASA Formal Methods*, R. Lee, S. Jha, A. Mavridou, and D. Giannakopoulou, Eds., Cham: Springer International Publishing, 2020, pp. 353–360.

[29] Scientific Working Group on Digital Evidence (SWGDE), "SWGDE best practices for mobile device forensic analysis," Best Pactices, Sep. 2020.

[30] Court of Justice of the EU, *C-548/21, C.G. v Bezirkshauptmannschaft Landeck*, ECLI:EU:C:2024:830, Oct. 2024.

[31] Supreme Court of the Netherlands, *Court ruling on the legality of investigating data carriers*, ECLI:NL:HR:2025:409, Mar. 2025.

Symposium: Digital Traces on the Scales of Justice

19 September 2025, Open Universiteit, Heerlen, The Netherlands
Modern society is becoming increasingly digitized. Digital evidence is therefore playing an increasingly important role in criminal investigations and court cases. Examples include data from mobile phones, emails, social media messages, GPS data, and online transactions. This data is often extensive and complex. Careful interpretation of such "digital traces" is crucial in order to be able to use them as evidence in a criminal case. Prior to the inaugural lecture, this symposium highlighted the importance of digital evidence and the complex legal frameworks for digital investigation. It also explained how the application of formal methods can help to better interpret the available digital traces.

Opening
Drs. J. (Jeroen) Venema, chairperson
product manager Hansken, Nederlands Forensic Institute

Between fiction and reality. What does the criminal lawyer of the future understand about digital technology?
Mr. M.M. (Martijn) Egberts
National Public Prosecutor for Digital Investigation,
Public Prosecution Service of the Netherlands

Lightning talks
Computer Science department, Open Universiteit

On digital evidence in criminal cases: trust is good, control is better
Prof. mr. dr. W.H.B. (Wilma) Dreissen
Professor of Criminal Law and Criminal Procedure, Open Universiteit

Lightning talks
Computer Science department, Open Universiteit

Formele Methods for Digital Forensics
Em. prof. dr. J.C.M. (Jos) Baeten
fellow of Center for Mathematics and Computer Science (CWI)

We all continuously leave behind digital traces. If we want to use these traces for a criminal investigation, it is more difficult than it seems. What exactly do they mean, and how can we be sure?

The Chair in Digital Forensics at the Open Universiteit aims to work with the Netherlands Forensic Institute to find answers by interpreting digital traces more accurately within legal frameworks, partly through formal methods, in order to contribute to a stronger criminal justice chain in the digital age.

Harm van Beek has focused on the use of digital data as evidence in criminal cases since 2009. He does this as a senior scientific researcher and forensic expert at the Netherlands Forensic Institute and, since September 2024, also as professor of Digital Forensics at the Open Universiteit. Harm obtained his PhD in formal methods (technical computer science) at the Eindhoven University of Technology (2005). Previously, he was co-founder and CTO of ISAAC, now part of iO, a company focusing on digital strategy and the design, development, and integration of digital systems.

Open Universiteit