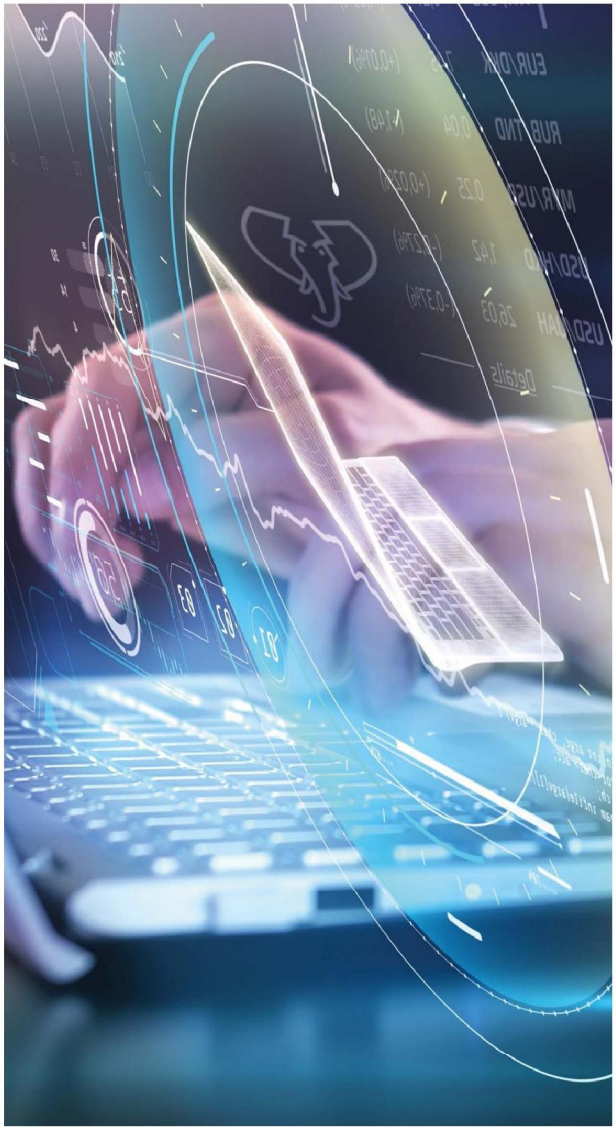

Hansken
The open digital forensic platform
Solution Paper &
Best practices
2025 EDITION

# Contents


| | |
|---|---|
| Welcome to Hansken | 3 |
| Solution Overview & Functionality | 4 |
| User interfaces and search tools | 6 |
| Joint innovation with Hansken | 8 |
| Solid forensic foundation | 9 |
| Hansken deployment, support and community | 10 |
| Hansken Functional Overview | 12 |
| Best practice | 14 |

# Welcome to Hansken

**Hansken is a digital forensics platform for fast filtering, searching, and analyzing millions of digital traces from seized computers, smartphones and several other devices. Thanks to an intuitive user interfaces, case investigators can search digital data without specialist knowledge, while Hansken also offers numerous specialized tools for more advanced and complex analyses. This way, Hansken facilitates collaboration between case investigators, digital experts and data scientists. Hansken streamlines digital forensic investigation and makes it more targeted, effective, and efficient.**



## Proven, scalable big data forensic platform

The Hansken platform consists of an advanced extraction engine, a specialized trace database, and extensive search and reporting capabilities. With proven components such as Hadoop and Elastic Search, a scalable, high-performance big data platform has been created which can search terabytes of data in real-time. The platform is ideally suited to finding and correlating relevant information within massive amounts of data. Successful cases vary from significant bank fraud cases and money laundering constructions to environmental crimes. Hansken also plays a crucial role in intelligence investigations and searching crypto communication servers.

## Meets judicial requirements

Each investigative service can run its own Hansken platform and control the forensic data in its own domain. The solution has delivered digital evidence and forensic substantiation in thousands of cases. Data are processed transparently and traces can always be traced back to the seized devices. Hansken thus registers and monitors both the chain of evidence and the chain of custody from crime scene to court. All user actions are logged unequivocally and can be retrieved by security and audit systems. Also, Hansken supports inspection and reporting functionality for defense lawyers and the judiciary.

## Adaptive to forensic developments

In our 'digital law enforcement race,' Hansken can quickly implement new requirements and wishes. As an integration platform, we use leading third-party core components and focus our own developments on specialist forensic search, reporting and (AI-based) analysis innovations. Our teams of analysts, developers and specialists develop new functionalities in a mature Agile and DevOps setting with a short time to market. Via the available APIs and the associated Software Development Kit, customers and partners can develop their own plugins and use them directly within Hansken.

## Professional support, strong law enforcement community

The Hansken platform is developed and maintained by the Netherlands Forensic Institute (NFI), one of the world's leading forensic laboratories, and is only delivered to law enforcement and intelligence agencies. Full support, documentation and online training are available. Via the Hansken Community, partners can exchange knowledge and jointly develop, share and deploy plug-ins. This way, law enforcement agencies bundle their expertise and join forces to boost digital forensics. Hansken is all about powerful digital investigation, leading innovation and sharing forensic knowledge.



75008

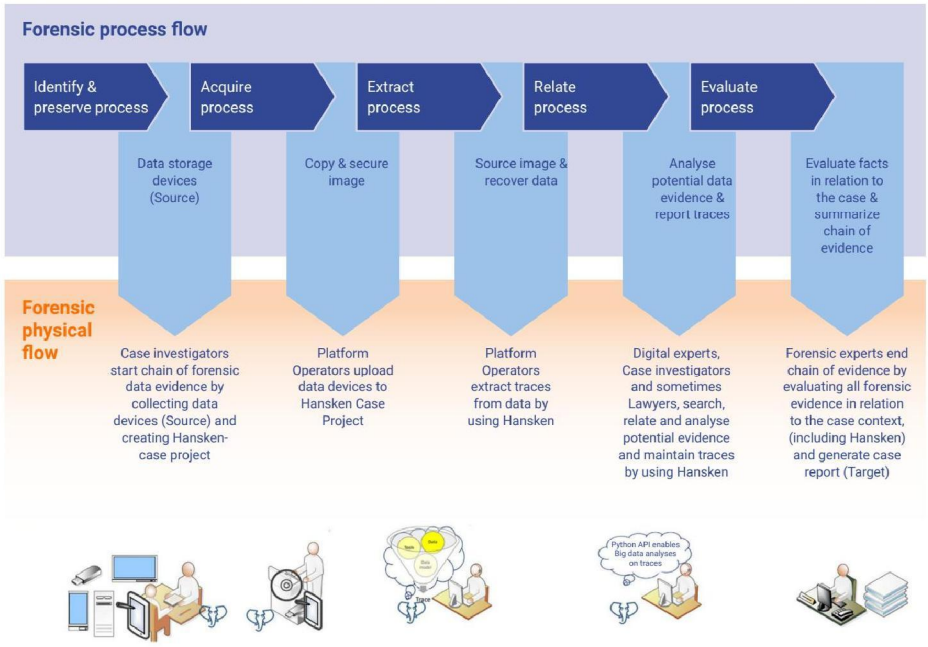
# Solution Overview & Functionality

**Hansken is a platform that offers Digital Forensic as a Service (DFaaS). This means that the forensic process for digital material is supported by Hansken DFaaS. The forensic work that involves solving criminal cases ranges from recognizing and preserving relevant (crime related) traces at crime scenes to explaining forensic results as expert witness in court. The forensic work referred to as Forensic physical flow follows the Forensic process flow. Both flows are visualized on the right.**

Hansken helps case investigators and digital specialists to access, analyze and evaluate digital material in the context of criminal justice. It is a modular platform where functions are available via APIs and user interfaces are decoupled from the core platform.

## Case administration and image store

Hansken can process data from devices like smartphones, laptops, wiretaps, and cloud storage. The device data is acquired and stored in a forensic Image Store using storage solutions like HDFS, CEPH or a local file system. Hansken also provides a Case Administration database to register cases, a list of images, image metadata and other case-specific log information.



## Extraction Engine

A core element is the advanced Extraction Engine, which produces an extraction of digital traces from the stored images. The engine builds on the Hadoop MapReduce framework and typically runs dozens of native and 3rd party forensic tools iteratively. All tools have their own function in this process (e.g., extract emails from email databases, extract camera information from photos, determine keywords, or calculate file attributes). Hansken maintains an extensive trace model to structure all these trace snippets into a comprehensive set of traces. The platform supports anything from documents and pictures to chat messages and system logs. The platform collects properties, metadata, and a full keyword index for each trace. The platform uses intelligent algorithms to classify and enrich traces and identify relations between them.

Extractions can be tailored and optimized to a specific case and its characteristics. By default, a pre-configured set of tools is defined for each extraction, but if appropriate individual forensic tools can be turned on or off on request. Also, the extraction process – or parts of it - can be repeated with other settings or with new or different tools. In this way, previous extractions can be replaced or supplemented, for example, when new or renewed trace formats are encountered during the investigation.

## Trace Database

Per trace, the metadata and a keyword index are stored in the Trace Database. Virtualization is used for the trace data itself, describing the data as a transformation of the original image rather than copying and storing them separately. Besides traces extracted from forensic images, digital experts can also insert external traces in Hansken from other sources. Such imported traces are included in Hansken's search engine.

## Filtering and Searching

Using the leading Elasticsearch technology, all traces are accessible and searchable for investigation. Traces in the trace database can be searched manually via user interfaces or automatically using scripts. Searching can be done in several ways. Using the metadata, it is possible to filter - for example – photos taken with an iPhone, e-mails from one specific sender, or documents printed in a certain period. Using the keyword index, it is possible to search for traces containing specific text strings. By combining multiple search methods, advanced search queries can be constructed. During the investigation process, traces can be inspected in detail, and users can add notes and labels per trace for investigative purposes. Important traces can be collected in so-called evidence containers, and reports can be generated.

## API's and User Interfaces

Via the Hansken gatekeeper – a RESTful API – we serve all operational and investigative functions available in Hansken. Multiple graphical user interfaces are available on top of Hansken and a scripting API is supported to simply write Python scripts.





75008

# User interfaces and search tools

Hansken is not intended as a substitute for existing forensic tools. On the contrary, the platform integrates and correlates the output of such tools and makes them in a user-friendly way manageable, accessible, and searchable for specialists and investigators. It offers an essential improvement by streamlining the digital forensics process.

Traditionally, digital experts first collect the digital traces, and subsequently, case investigators try to use this information to solve the case. The scarcity of specialists is here often a bottleneck. With Hansken, investigators can search the data themselves and immediately benefit from their specific case knowledge. Digital experts can focus on the more specialized and challenging analysis while they also coach the case investigators. Hansken supports this with multiple user and expert interface, and other specialist and scripting tools. Via APIs, these tools access the same core functions and data of the Hansken platform.

## Tactical User Interface

Using the Tactical User Interface, a case investigator can – without specialist knowledge - filter and search the digital traces on file properties, metadata, time periods, geographical areas, and keywords. Via an intuitive navigation menu, users have access to an extensive set of pre-configured search filters that can easily be adapted to the investigator's wish. You can further refine the search results and highlight relevant traces. The investigator can also analyze the details of a trace, add tags and annotations, and create so-called evidence containers and reports. The GUI facilitates cooperation within investigation teams.



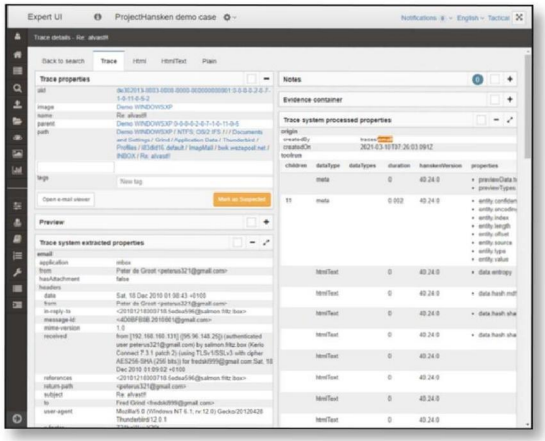
## Advanced capabilities

For experienced investigators and digital specialists, Hansken offers more comprehensive tools to search and analyze the search data. For example, the Hansken Query Language (HQL) makes it possible to compose complex filter and search queries based on the extensive trace model and all trace properties. Moreover, it is possible to combine multiple individual queries into a compound search query, while Hansken also offers extensive scripting options to automate searches.

## Technical User Interface & API

There is also a technical interface for experienced investigators and specialists. Through this interface, you not only have access to the data that the Tactical GUI uses, but you also have access to a lot of underlying technical data and tools, extensive visualization options, and statistics. Users can easily switch between the tactical and technical GUI. Thanks to these different interfaces, every investigator and specialist can simultaneously work on the same case from their own perspective. Hansken thus stimulates cooperation, making digital forensics much more efficient and effective.

For administrators the technical user interface is to start and monitor extractions. For expert users and developers, Hansken offers a complete API to access data using scripts, like Python.




**Technical interface** (experts/administrators)

```python
#!/usr/bin/env python3
# encoding=utf-8

from hansken.tool import run

# we define a function to do the things we want to do
def search_and_process(context):
    # our argument "context" is a hansken.remote.ProjectContext, see the docs
    # for what it can do other than searching for traces
    with context:
        # query was copied from Hansken Expert UI
        query = ''' type:file '''
        # search for traces matching the query
        results = context.search(query)

        for trace in results:
            # process the results, print some details for each trace we've found
            print(trace.uid, trace.name)


if __name__ == '__main__':
    # call the hansken.py command line, but make it call our function
    run(with_context=search_and_process,
        # the gatekeeper REST endpoint when this script was exported, note that
        # this can be overridden by passing -e/--endpoint on the command line
        endpoint='http://localhost:9091/gatekeeper/',
        # the keystore REST endpoint when this script was exported, note that
        # this can be overridden with --keystore
        keystore='http://localhost:9090/keystore/',
        # the project id of the project named "Bad Guy",
        project='48619f8a-1e07-445b-ae0e-905392bc7400')
```

**Scripting API** (developers)

# Joint innovation with Hansken

Besides structural, architectural roadmap investments in, among others, API development and analysis functionality, Hansken is also designed to implement new requirements and trace types with a short time to market. New applications are constantly becoming available on computers and smartphones, and more and more other devices are also generating data, from vehicles and household appliances to sportswear. Also, data from applications that are only popular in specific countries, regions, or within particular communities may be relevant for forensic investigations.

## Agile and open platform

To implement new functionality with short lead times, Hansken uses a microservices-based architecture and an agile-based development approach. Essential is that our system and policies safeguard forensic quality, security, and privacy. The development and use of Hansken are always reproducible, and changes to the platform always uphold existing investigation results.

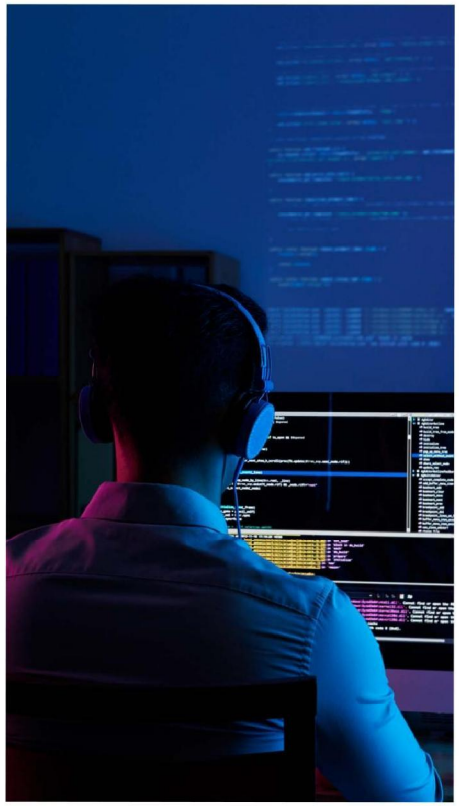
Moreover, Hansken is designed as an open system in which both Hansken and community partners can develop and insert functionalities as plugins. This guarantees a stable development path for the core Hansken framework, while facilitating the introduction of new trace types, analytical functions, and other features whenever necessary. Hansken plugins are a way to leverage the knowledge and expertise of our community partners.

## Powerful plugin SDK

An important tool is the extraction plugin Software Development Kit (SDK), which facilitates the rapid development of new plugins. The SDK is well-documented, supports commonly used programming languages such as Java and Python, and developers can create their plugins with a short learning curve. In addition to new digital formats and trace types, like e.g., cryptocurrency wallets, the SDK also facilitates introducing more advanced trace analysis by combining existing functions and (AI) technologies. Using such advanced plugins, we can verify whether video materials are only stored on a device or actually played, automatically recognize the use of specific languages in documents, create speech-to-text extractions from audio files and search for particular objects in images, to name just a few options.

## Joint development

More and more parts of Hansken are open for joint development. Joined development means that the source code is available for Hansken partners and everyone with access is able to make changes to the product, supervised by a team of experts. Most important element is the Tactical User Interface (a.k.a. HUIB), which has been developed by one of the partners. It is now open for joint development and already four partners are working on its shared code base. Translation of the interface are done by international partners, close attached to the domain knowledge of their own country. More and more parts of Hansken will be available for joint development, including parts of the Hansken core and of development kit.





75008

# Solid forensic foundation

Over many years Hansken has been used in hundreds of cases. In high-profile lawsuits, detailed questions have been posed about the use of the search engine, data entry and editing, accuracy, and quality control. Hansken withstands judicial review and has clearly shown that Hansken is essentially a tool to help search through large quantities of data and has no influence whatsoever on the information found.

All data are processed transparently, and all reported traces can be traced back to the source (e.g., the seized telephone or computer). Hansken provides safeguards for the correct processing of the evidence itself (chain of evidence) and a proper recording of this investigation process (chain of custody). Hansken also supports the correct handling of privileged communications, and case data can be made accessible to lawyers, prosecutors, and the court. Hansken facilitates a solid separation between system administrator roles (infrastructure, hardware, underlying software, and Hansken platform) and functional administrator roles (use of Hansken for case investigations). All actions and requests to Hansken, including the date, time, and request details, can be routed to a Security Information and Event Management (SIEM) or other external audit systems.

## Evidence handling

Source data from seized devices are secured in evidence files, also called forensic images. Such an evidence file is typically produced by the staff of the investigative service and stored outside Hansken. An evidence file contains information about the securing process, and hashed digests are calculated to ensure the integrity of the source data.

For the extraction, copies of the evidence files are made. Such a copy is compressed, encrypted, and stored in a Hansken-specific format for fast and secure processing. This conversion does not change the source data itself; any file attributes of the source data are also copied. Hansken provides the ability to check the file digests of the copied data after it has been placed in Hansken. For each evidence file, Hansken records its unique ID, a timestamp, the responsible administrator, and the file characteristics and digests of both the original and the copied file.

## Extraction handling

All settings, actions, and results are logged for the entire extraction process and the individual trace extractions. The same happens for any manual insertion of additional searchable data. A unique ID and link to its source data are listed for each trace, a digest is calculated, and the used forensic tools and version numbers are captured. For the overall extraction process, Hansken records the start date, time, extraction settings, statistics, and all forensic tool settings. Trace errors due to, e.g., corrupted files or changed file format, are caught and captured.

## Searching and reporting

Typically, filtering, searching, and global inspection are done using the metadata and keyword index in the trace database. The trace data cannot be modified, but users can add notes and labels for investigative purposes. If a trace seems relevant for the investigation, a user can via a link in the trace database view or download the source data in the original evidence file. Hansken plays an essential role in searching traces but then always refers back to the source data for final trace evidence.





75008

# Hansken deployment, support and community

**Hansken is only delivered to law enforcement and intelligence agencies, to which we offer a smooth adoption model and a scalable growth path. Hansken is a non-profit initiative, and a fair license model funds innovation, development, maintenance, and professional support. In addition, partners are invited to join the Hansken community, where they work together on the further development of Hansken.**

## Adoption model, from demo to roll-out

Hansken offers an adoption model in which partners can gradually get acquainted with the platform and step-by-step migrate toward a production environment. Potential partners often start with an NFI demo account. A logical next step is a pilot period with test data at the NFI pilot platform. From there, a logical follow-up is an on-premises pilot environment with own production case data. Based on the pilot results, the partner can decide on a full Hansken license and roll out their own Hansken platform in their infrastructure.

## A scalable growth path

Hansken is a big data platform with unique scaling capabilities. The platform can be scaled up with the actual use in the organization and support over a hundred concurrent cases, with over 1000 seized devices per case, more than 100 terabytes of data, and hundreds of millions of traces. Implemented with a relational database and multiple Hadoop and Elasticsearch clusters, a Hansken partner's private cloud environment can be tuned exactly to the partner's requirements in terms of authentication, authorization, data encryption, users, and performance. Also, a containerized version has been developed, simplifying the roll-out, making the architecture more flexible, and automating maintenance and continuous improvement.

Hansken is an open platform, it has the ability grow and adapt to an organization's need. It allows for automation of the process, tailor made adaption and features. Mature Hansken partners have a "devops team" surrounding the platform to create a perfect fit for their organization.

## Hansken Training and Documentation

The Hansken Academy offers an extensive set of courses for different types of users and learning paths. The Hansken Fundamentals and Intermediate courses have been developed for investigators, while advanced and specialized courses are available for experienced investigators and specialists. The training courses are hands-on and use fictional cases which students can practice in real life. Many courses can be followed individually and online. In addition, Hansken comes with extensive documentation, integrated as much as possible into the Hansken user interfaces.

## Professional support

The support staff of the Hansken partner itself can do 1st and 2nd line support to their end-users. Hansken provides all necessary training and documentation for the partner's team. Hansken runs a professional support organization for 3rd and 4th line requests that handle more technical and complex issues, including questions related to 3rd party products. Partner administrators have direct access to the Hansken portal and support desk.

## Join the Hansken community

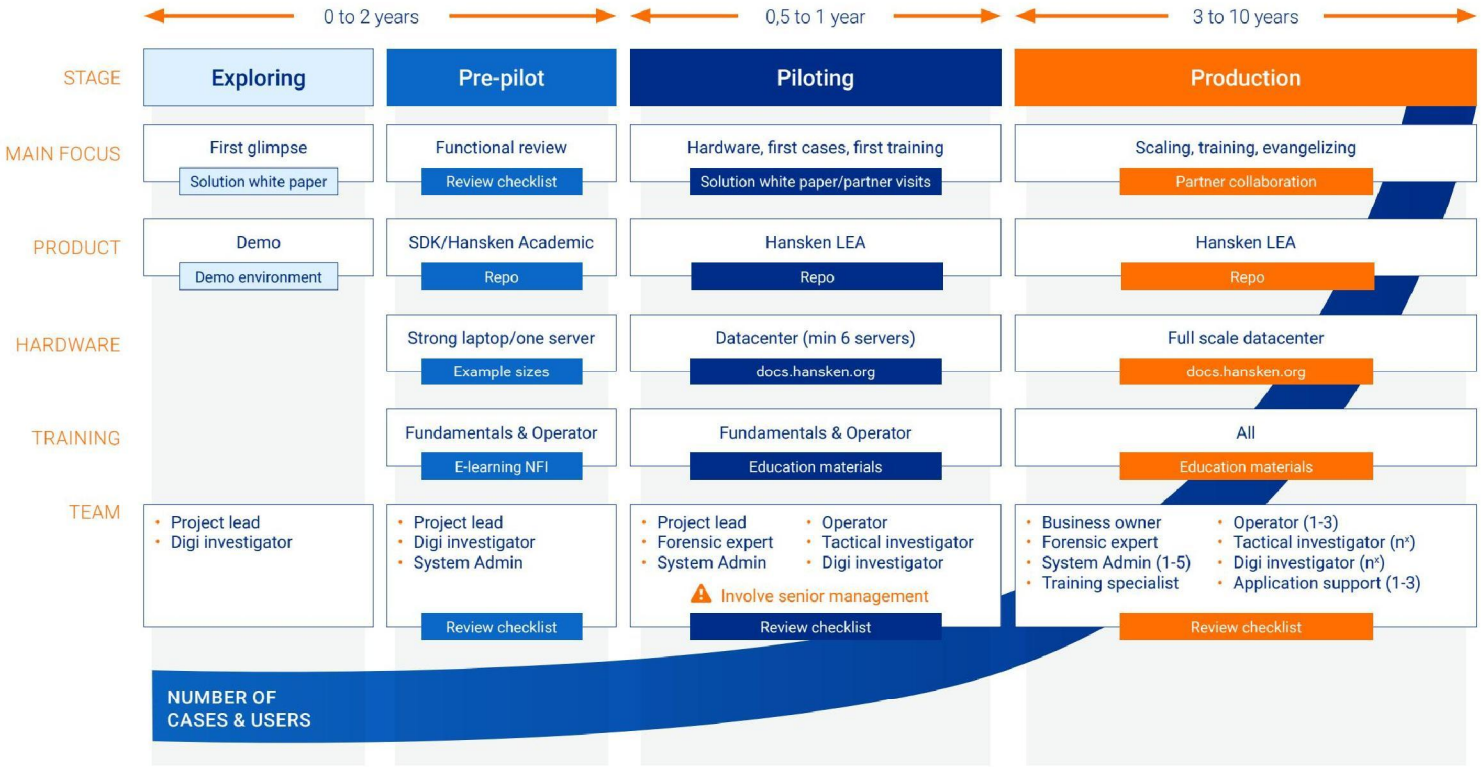
Hansken partners can join the Hansken Community to bundle their expertise and join forces to deepen digital forensic knowledge. The Community is for and by law enforcement and intelligence agencies and collaborates with the Public Prosecution Services and the Digital Forensic academic community.

The Hansken Community provides an open platform where software can be developed, shared and deployed immediately. There are sub-communities where members collaborate on specific plug-ins. These are available in a shared Hansken app store and can be used directly in ongoing cases by all members.

The Community offers a Portal where users can connect and exchange information. The Portal also offers Hansken Academy materials and links to the Hansken practicing environment. These facilities provide new users with quick, practical training and support experienced users in keeping their digital skills up to date. Community Days are held frequently to share Hansken-related experiences and new insights.



75008


0 to 2 years
0,5 to 1 year
3 to 10 years
STAGE
MAIN FOCUS
PRODUCT
HARDWARE
TRAINING
TEAM
Exploring
Pre-pilot
Piloting
Production
First glimpse
Solution white paper
Functional review
Review checklist
Hardware, first cases, first training
Solution white paper/partner visits
Scaling, training, evangelizing
Partner collaboration
Demo
Demo environment
SDK/Hansken Academic
Repo
Hansken LEA
Repo
Hansken LEA
Repo
Strong laptop/one server
Example sizes
Datacenter (min 6 servers)
docs.hansken.org
Full scale datacenter
docs.hansken.org
Fundamentals & Operator
E-learning NFI
Fundamentals & Operator
Education materials
All
Education materials
Project lead
Digi investigator
Project lead
Digi investigator
System Admin
Review checklist
Project lead
Forensic expert
System Admin
Operator
Tactical investigator
Digi investigator
Involve senior management
Review checklist
Business owner
Forensic expert
System Admin (1-5)
Training specialist
Operator (1-3)
Tactical investigator (n*)
Digi investigator (n*)
Application support (1-3)
Review checklist
NUMBER OF
CASES & USERS



75008

# Hansken Functional Overview

**last revised release: 47.12.0**

### Versions

- Academic version
- LEA version (same as Academic version, but includes proprietary or licensed extraction tools)

### User interfaces

- Hansken UI from BZK (HUIB)
- Technical UI (Expert UI)
  - Operator Functions
  - Documentation
  - Notebook support for Python-based extension

### Operations

- Create projects and images
- Upload image data
- Link images to projects
- Configure, schedule (pause and continue) and run extractions
- Manage own projects and images
- (Partial) cloning of indexes
- Monitoring of tasks and health, including requests to underlying systems (e.g., Elasticsearch and Cassandra)
- Backup and restore projects (tags, notes, user-added traces and project metadata and project images metadata)
- User permission management
- Portable case support: generate a sliced image (e.g., to support removing op privileged communication)

### Extraction engine

- Extract to new index or append to existing index
- Partial extractions, based on query for selecting traces to process
- (partial) index cloning
- Select which registries and event logs must be processed
- Support for (Dockerized) extraction plugins
- Support for VNFI images (local/network folder with files can be processed as it was an image)

### Extractions

- Hash calculation (MD5, SHA-1, SHA-256) for all traces
- Matching against the hash databases, e.g. NSRL and HVEICSE
- Automated decryption of PGP and S/MIME messages and EsoCrypt encrypted pictures if a key is available
- Carving support for several structures (e.g., office documents, pictures, PDF document, MP4 movies, SQLite)
- Generic representation of database tables as comma separated values
- Add binary objects stored in databases tables as child traces to the tables
- Support for several images, including UFDR reports, wiretaps (with location details) and TomTom navigation systems
- Support for multipart ZIP files
- Support for specific data structures from cryptophones (including Ennetcom and Encrochat)
- Support for storing AI models and for configuring tools to use these models
- Support for third-party tools (GoogLeNetCaffe, Protorec, Tesseract, Sleuthkit for APFS)
- Traces are built from files and metadata in UFED ZIP archives
- Chain of custory maintained, including successfully run and failed tools per trace.
- Image hash calculation and validation
- File system support for BTRFS, QNX (FFSv3), XFS
- Support for ESEs, including (corrupt) transaction logs
- Single files can be extracted via the Rest API



75008

- Basic analyses of wiretapped internet information in common version of the PCAP file-format
- XML-representation of binary plist files
- Support for PhotoDNA (LEA version only)
- ZIP bomb detection (for types 'boom' and 'overlap')

### Trace enrichment

- Classify audio as speech or music
- OCR of pictures including PDF files
- Recognition of faces (including age classification), bar codes and scene texts
- Classify pictures and video frames based on a classification model (FIRE), including money, knives, soft drugs, hard drugs, weapons and porn
- Videos can be previewed as thumbnail film strip
- Number of recognized faces available as misc property to pictures
- Number of pages available as misc property to documents
- Human readable text streams are used for previewing traces
- Relations between different traces in the trace model
- All audio can be transcoded to WAV

### Trace model

- Support for traces (e.g., emails, files, pictures, chat messages, documents)
- Support for tracelets (e.g., entities like email addresses, urls, ibans and crypto currencies) in trace data or metadata
- Vectors support (e.g., for face representation)
- Store tool run details
- Store AI-based predictions (including classifications)
- Store user-added traces
- Property for uniquely identifying traces

### Trace querying

- Searching and filtering of traces as well as tracelets
- Facet support (aggregate result)
- Retrieve all unique values for a property
- Search across multiple projects
- Search for emoticons
- Search by regular expressions
- Search with (leading) wildcards
- Support for search suggestion, including Arabic/ Greek/Russian character sets
- Timeline search
- Search for geo location
- Search for vectors (based on distance)
- Deduplicate search results based on configurable trace property
- Traces can be downloaded for offline handling
- New traces can be added by end users

### User presentation

- Autofill suggestions for both queries and keywords
- Gallery shows first page of PDF documents as thumbnail
- Preview texts for traces text contents
- Keyword search hits are shown in their context (surrounding text)
- Search hits can be highlighted in data streams
- Support for all character sets, including Arab, Chinese and Cyrillic
- Show media captions in chats in preview

### Documentation

- Functional release notes
- Technical release notes
- API documentation
  - REST API
  - Python API
- Extraction plugins
  - Extraction plugin SDK
  - Extraction tools
- User guide:
  - Hansken Query Language (HQL)
    - HQL human
    - HQL json
  - System administration
    - Backup and restore
    - Cassandra to Postgres migration
    - Elasticsearch Shard Management
  - Traces management
    - Export to kafka
    - Identities
    - UFED supported model types
    - Quality objects
  - Operational demands & context
    - Privacy and Security measures
    - Forensic measures
  - Virtual NFI image format



75008

# Best practice

## Introduction

This best practice has been written to give insight in a successful implementation approach of Hansken. This document describes the current solution of Hansken at one of our partners. This implementation is one of the largest with cases up to 700 devices and individual cases of 20+ terabytes and might give relevant insights in the way to optimize the use of Hansken.

## General approach

Implementing Hansken in the right manner, starts with understanding the goals you want to achieve with the implementation. What is your organization type? Who will be your end users? Which departments will be involved and impacted? How does Hansken fit into the process and where do I need to make changes?

The partner has identified 4 end user groups:

- Operators or maintainers of the platform
- Tactical investigators or analysts
- Digital investigators
- Developers

By involving departments and users in an early stage and finding success stories that link back gains a positive attitude towards Hansken within any organization.

There always is an extra need for using clean data in the platform, since companies have very strict security demands. Hansken has added and integrated the cleaner service in order to achieve this, based on the experiences. If you are working with sensitive materials in a restricted environment, this cleaning might help you convince your security officers.

## Lessons learned

The most valuable lessons learned from the implementation are:

LESSON 0

**Form a Hansken team and be ready to scale this team as well**

A Hansken implementation can start with a small project team. Having a dedicated and multidisciplinary team helps to overcome hurdles at the start. But it is also important to note that this team will scale along with Hansken en keeping a dedicated team will help you during the use and scaling of Hansken.

LESSON 1

**Find ways to have impact on network design**

Multiple network segments are used for security reasons. This increased the need to have direct access to network specialists and have a direct influence on network plans. In most organizations network architects and specialists are hard to reach and involve. For example, Hadoop and the Hansken services both have a specific network segment, but communication between the segments is done without using firewalls, to ensure best performance.

It is important to influence the network design around Hansken, to be able to handle data from 'unclean' segments to 'clean' segments, meaning the infrastructure offers different levels of security within the organization. This is especially important in highly classified environments.

LESSON 2

**Scaling is a learning game**

Although scaling is by design in Hansken, it is important to notice that scaling the application is also a learning experience. It is difficult to predict which parts of Hansken needs to be scaled in what size and order. Scaling storage could result in the need to scale the trace or gateway services as well, depending on the demands that comes from the increased usage of Hansken. It is essential to keep overview of all Hansken services to monitor their health. Another lessons learned is that scaling on a larger scale than planned helps overcome risks. If you plan to add one server, consider adding two or three. This will provide you with extra room to scale other unexpected services.

LESSON 3

**Automate, automate, automate**

Hansken will have success stories if processing digital evidence is easy. Automation is essential and organization specific. Prepare to have a team in place that can help automate and learn from other partners experiences in automation. Please check the separate paragraph on automation later in this document.

LESSON 4

**Handle any data stream**

Hansken is able to handle both structured and unstructured data. Typical forensic data streams will be images, USB sticks, ZIPS and UFED reports. By having a team that is willing to invest time in supporting new (organization-specific) data streams, you will increase

Hansken adoption. Most big data platforms are able to process structured data, but a lot of forensic data is also unstructured. In larger organizations, this helps to distinguish Hansken from other platforms. Examples of data streams that impacted the decision to use Hansken in some specific cases is the ability to process attachments in chat messages (in addition to just the chat messages itself). Another example are evidence items that keep increasing ("dripping in), like data from webcrawlers. The Hansken team designed special vNFI images to be able to process these data streams.



LESSON 5

**Make room for innovation**

Make sure there is room for innovation. Not only scaling is a learning game. Developments in the world of forensics are fast paced. Hansken changes continuously, it gets new possibilities and when you make use of it, along comes the need to make updates yourself. Have a team of people with capacity to develop GUI add-ons to tailor visualization to your need, extraction plugins to support new formats, and update and fine tune the upload and extraction as new demands come along. 10% is a good rule of thumb to have available for innovation.



LESSON 6

**Find sponsors and ambassadors and mug them**

This might feel like a no brainer, but pinpointing the most important stakeholders in your organization helps. Sponsors, staff departments and end users must know that Hansken is there to help them. But you will also need budgets and room to grow. Try to find simple ways to involve them, keep them informed, but also 'Hanskenize' them by using a goodie or gadget like a simple Hansken Mug.

## Automation

The process of extracting data streams or digital evidence benefits from a high degree of automation. Create an automation tool that has 3 key important aspects:

1. A simple UI to browse all images in their own storage folders
2. A set of presets for the most important types of extraction
3. A simple "Play" button to start the extraction process

The automation tool uses backend scripts against the automation API from Hansken to do so. What really helps in the speed of processing is to have one service account that directly handles the upload to Hadoop. This mimics the working of a mail slot: the system account only has access to upload data to Hadoop.

### Initial and append extractions

By default the automation tools handles the initial extraction of the data, but when the extraction is successful, the scripts also automatically start append extraction for more computation intensive tools, like OCR, translating and nominatim. These are time consuming extraction tasks that can easily be done afterwards using append extractions.

### Unsuccessful extractions

When extraction is not successful, there needs to be manual labor to find out what went wrong. More insight in progress (percentage complete or estimated hours to go) is still needed to improve this and requests have been made to the Hansken Development teams to achieve this.

Partners created custom process pages to continuously see the status of the running extractions per case. This is a custom made function that polls backend processes, and displays the actual extraction status. It also supports the priority queuing line that has been custom made as well. This is in place to be able to prioritize urgent case extractions.

When extraction fails, a notification is pushed towards the internal (mattermost like) communication application, in order to warn operators in time. Knowledge on well known failures is still gathered to be able to respond even quicker to failed extractions and find ways to automate responses as well.

### Deleting projects & images

Currently all projects and images are staying in Hansken for archiving reasons. There are some cases in which deleting of materials is required by law. This has been fully automated as well. Expiration dates of projects have been stored and users will receive notification before the end date has been reached. If there is no reaction, the deletion of these projects will be executed automatically as well. This is all custom automated as well.

### Some findings on hardware used

In addition to the hardware requirements, here are some experiences after several years of running Hansken: Elastic should be running on SSD disks, which drastically improves query performance. Elastic has had storage issues in the past. The lessons is to continuously monitor storage, cpu and memory usage. The (new) cleaner service drastically increases the storage of Cassandra (soon to be S3), this is something to pay close attention to.



75008

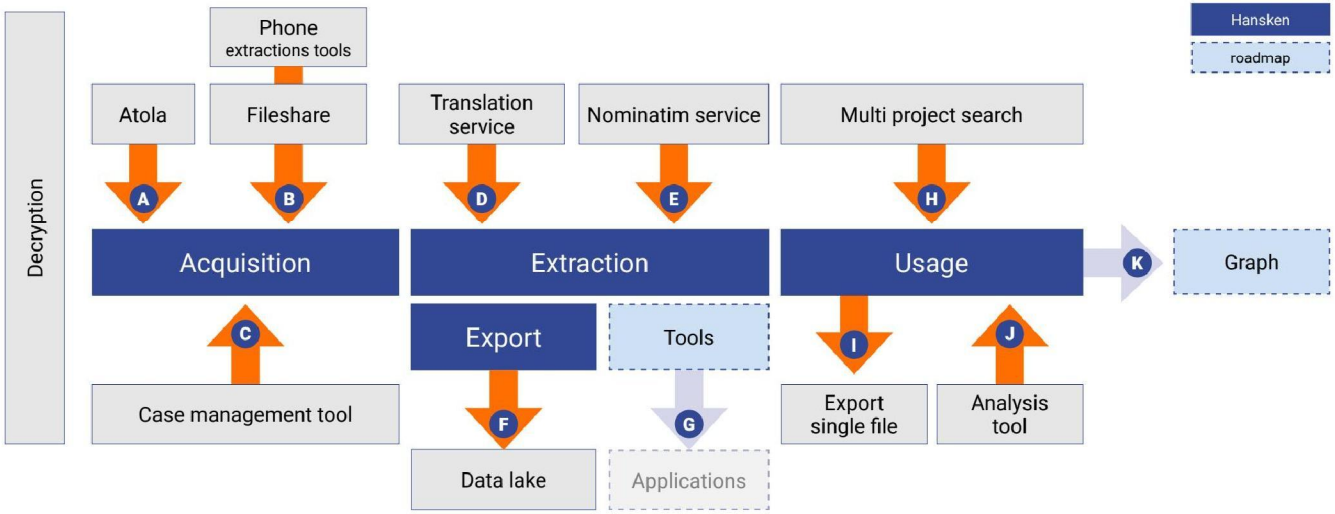
## Hansken ecosystem

Hansken is not an island. It resides in an ecosystem of applications and tools. The relevant items from a example install are depicted below. Note the corresponding letters indicating the connection to Hansken, they are each explained below.

Important to note that three main stages are used: acquisition, extraction and usage. The acquisition step in this example is by default not handled in Hansken, but a small portion of the acquisition work is handled by the automation tooling before it starts the extraction stage.

The following diagram shows an ecosystem, showing which applications, tools and processes interact with Hansken.

Figure 1: Typical Hansken ecosystem




- A Evidence items processed by Atola
- B Data streams from file share
- C Case management information
- D Translation services
- E Nominatim services
- F Export to data lake / data ware house
- G Tools as a service (roadmap item)
- H Multi project search
- I Single file exports
- J Analysis tooling
- K Graph visualization (roadmap item)

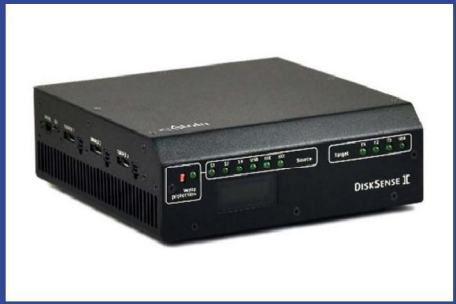


75008

**A - Images processed by Atola**

This example uses the Atola Insight Forensic, which is a fast forensic imager with the capacity to perform 3 simultaneous imaging sessions on a wide range of media.

It also offers complex yet highly automated data recovery functions on failing storage devices and provides utilities for accessing hard drives at the lowest level. The automation tool can handle data streams directly from the Atola, using its Web API. The Atola is used for every image that is not from a mobile phone.




**B - Data streams from file share**

Almost every forensic file that must be processed by Hansken is placed in a folder (per project/case) on a file share. The file share is accessible by the automation tool. Mobile phone images, typically processed with Cellebrite Forensic Workstations, are placed in the folder, but also single files and containers (like uncompressed ZIP archives) are in the file share folder. The automation tool has access to the folders and users can browse the folders to select evidence items that must be processed.

*Note that decryption is a completely separated step in the acquisition stage. Hansken is not able to handle this and there has been no need to so as of yet. The skills and tools for decryption are too different and specific.*
*If Bitlocker is used, it will be displayed as case alert in the Hansken UI.*

**C - Case management information**

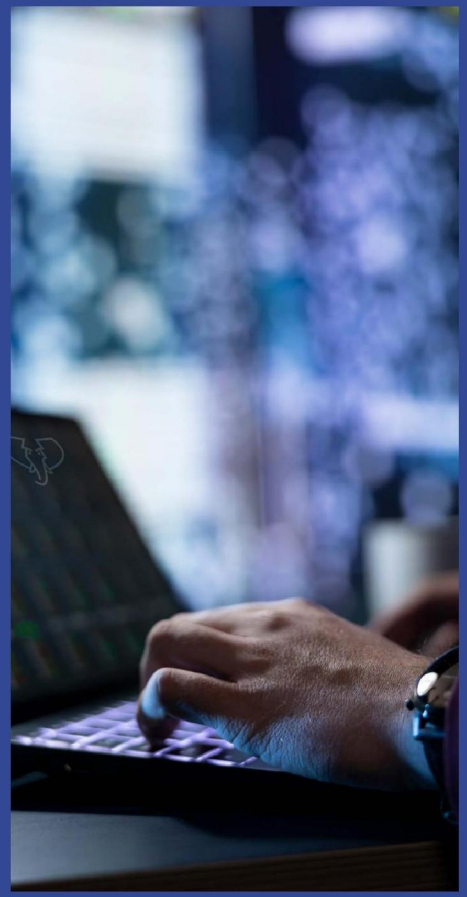
The projects and its specifics, including information on images and authorized users are stored in a separate case management application. The automation tool can access this information and use it to automate retention times, users and the location of the images that needs to be processed. Accessing the case management application saves a lot of time and prevents duplicate administration.

**D - Translation services**

Translation of evidence data is really important, since a lot of foreign languages are to be expected within the forensic data. Translation has to run as a separate service after the extraction, extracting only the content which needs translation, because it consumes a large amount of time. Translations are available in the front end and searchable within Hansken.

**E - Nominatim services**

Nominatim (from the Latin, 'by name') is a tool to search OSM (Open Street Map) data by name and address (geocoding) and to generate synthetic addresses of OSM points (reverse geocoding). It can be found at nominatim.openstreetmap.org. An on-premise service is recommended to be in place to enrich geocoding data with readable addresses.





75008



**F - Export to data lake / data ware house**

Various Data warehouses or data lakes use structured data from Hansken to enrich the data further for analysis purposes. Especially entities like chat and mail are exported to the data warehouse, using the export function in Hansken and placing the (JSON) data on a Kafka queue. The data is transformed from the Hansken data model to the Datawarehouse data model before it is stored in the data warehouse. This data mapping is done outside of Hansken.

The general experience is that a Datawarehouse can add much more enriched data and information to the evidence data (for example OSINT information), but is much less usable for ordinary end users. The functionality in and usability of the Hansken UI has a better fit to the needs of a larger amount of end users. At the intake of a new case, the question to the person responsible for the data is always whether export to the warehouse is necessary. Another aspect is that Hansken has more capabilities for preprocessing unstructured data, although Hansken is just one of the many sources for the data warehouse. As different types of data streams are still coming in, it is good practice to be able to process them with tools like Hansken and not bother the Datawarehouse with every type of data.

**G - Tools as a service (roadmap item)**

Hansken offers multiple tools to be used during extraction. Some of these would be interesting to use in other applications as well. In the long term, partners may be looking to expose multiple tools from Hansken, like OCR for example. This has not been implemented yet, but will be part of the eco system in the coming years.

**H - Multi project search**

One should create a tool that can access Hansken to be able to search across cases. In some circumstances this is allowed by law in the Netherlands. The tool accesses the Hansken API and has options to search and retrieve all results or just a hit or miss value (query has a result or not), depending on the mandate.

**I - Single file exports**

Some users are allowed to download (cleaned) evidence files to be able to process them further in other network segments. Downloading specific files directly from Hansken requires extra care in highly classified environments. The cleaner services help to make this step safe.

**J - Analysis tooling**

If further analysis has to be done on traces, Hansken data is available through the API for some analysis tooling. These tools can access the data or export it using the API and process it further. One of the examples is graph visualization as mentioned.

**K - Graph visualization (roadmap item)**

A Proof of Concept has been done to visualize relations between traces. This has been achieved through an export of a series of Hansken traces into a Neo4J database. This allowed the results to be displayed with Neo4J as related nodes in a Graph.



75008

## Get in touch!

Want to learn more or experience Hansken firsthand? Contact us at Hansken.Community@nfi.nl or visit www.hansken.org!

Hansken

The open digital forensic platform