

AI assisted investigations using Hansken

Harm van Beek

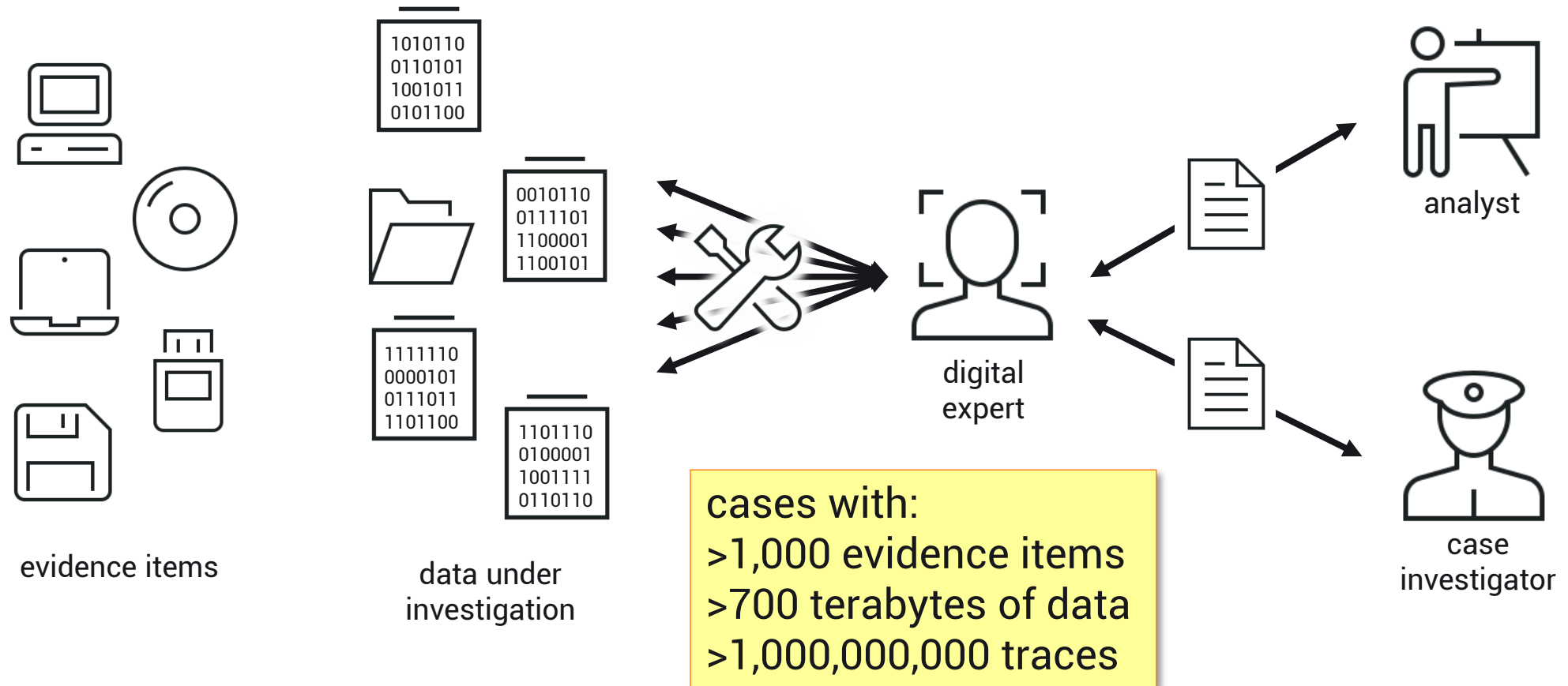
Netherlands Forensic Institute

CRIM/AI - Netherlands Country Roundtable

2 October 2023

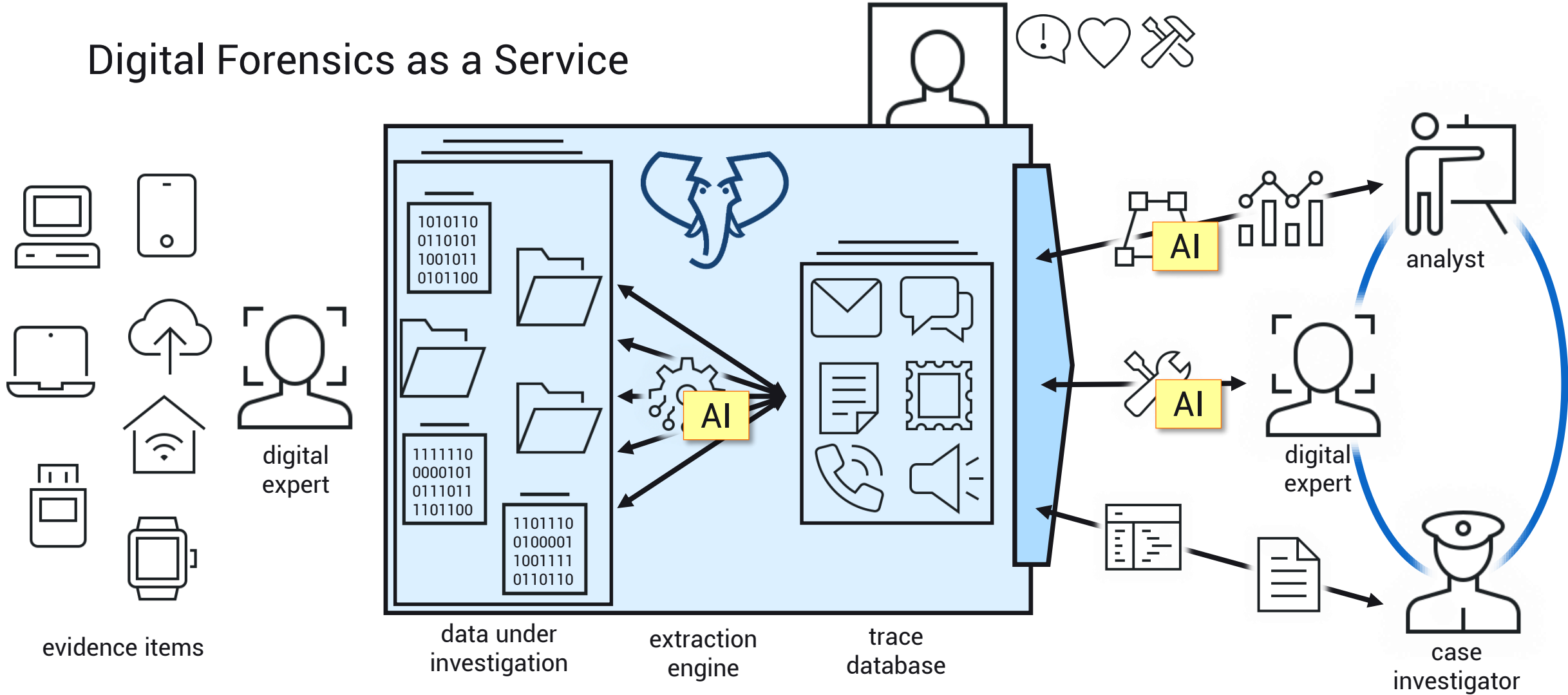


Digital Forensics (traditional)





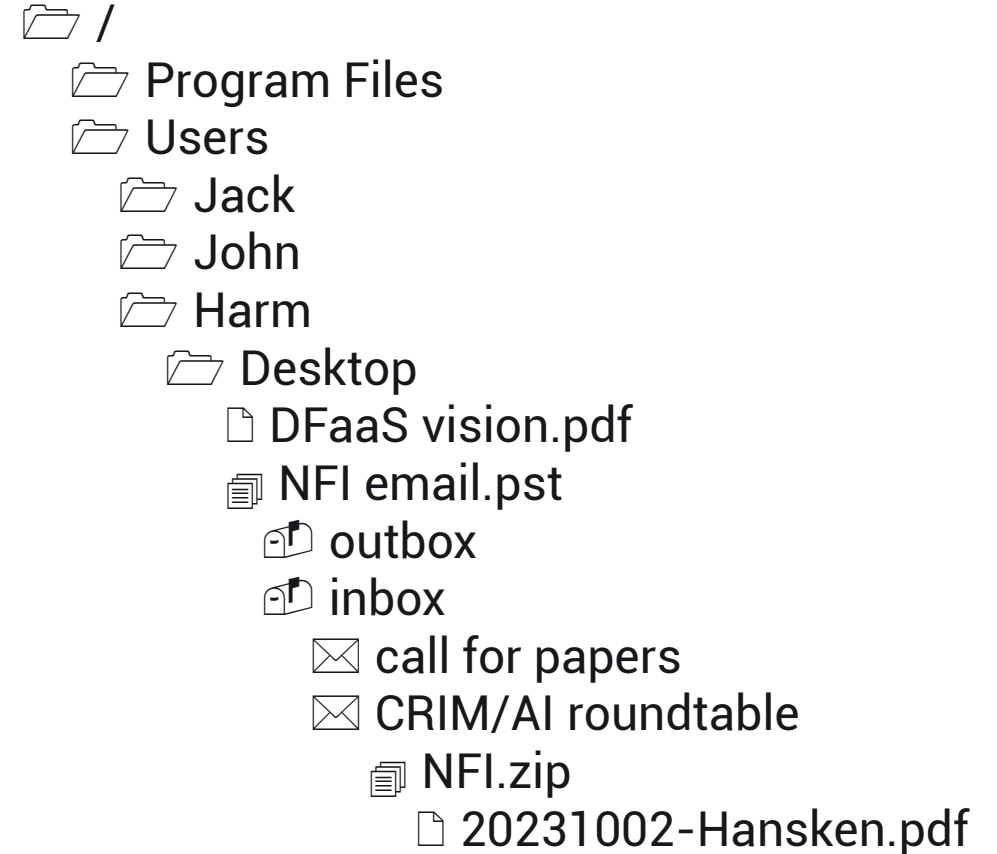
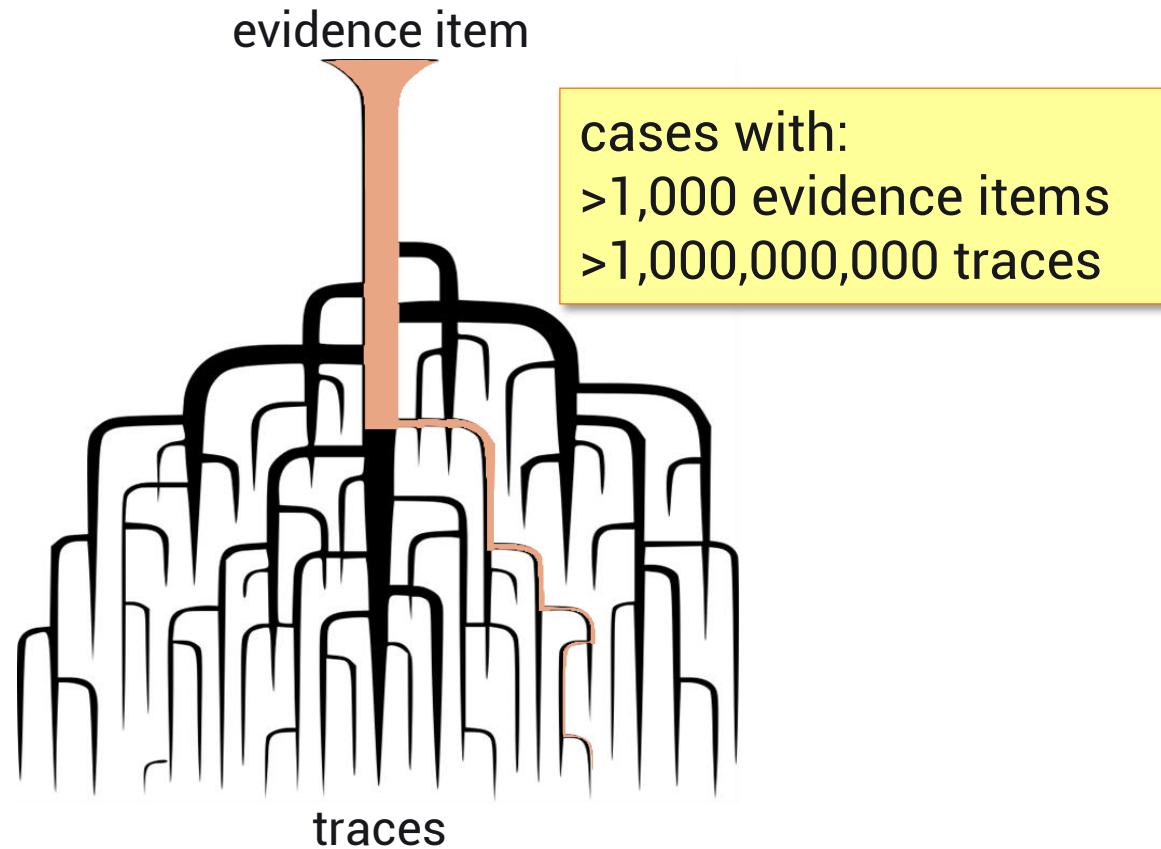
Digital Forensics as a Service





Trace normalization

so processing can be automated





Support investigators & experts

The image displays several screenshots of the Hansken software interface, illustrating its capabilities for investigators and experts.

Dashboard (Top Right): Shows a high-level overview of the system with various filters and statistics. The dashboard includes a search bar, a navigation menu, and several key metrics: Leads (3 items), CT relevant (5 items), Cyber relevant (0 items), RT relevant (0 items), and Pending (0 items). Below these are several charts and graphs representing different data categories like Communication, Documents, Database, and Virtual.

Search Results (Middle Left): Shows a search results page for a specific case. It includes a search bar, filters (Type of trace, Picture, Camera), and a list of traces. The results are displayed in a grid format, showing a list of traces with their respective images and details.

Trace Details (Middle Right): Shows a detailed view of a specific trace. The trace is identified as 'Re: alvast!!' and is of type 'email'. The details include trace properties (uid, image name, parent, path), notes, evidence container, trace system processed properties (origin, createdBy, createdOn, toolrun), and trace system extracted properties (application, from, hasAttachment, headers, received, references, return-path, subject, to, user-agent).

Code Editor (Bottom Right): Shows a Python script for interacting with the Hansken API. The script defines a function `search_and_process(context)` that searches for traces matching a query and processes the results. The script also includes a `main` function that calls the `search_and_process` function with the appropriate context.

Anonymous ▾

Documentation - Trace model

Trace model for the current project

A trace is a structured data object. The object contains intrinsic properties, which are mandatory properties for every trace, such as `name` and `id`. Furthermore, a trace can contain origins like `system` or `user`. This origin describes where the data is coming from. An origin contains categories, like `extracted` or `processed`. Every category has a list of allowed types, which are listed below. Every type contains properties, for example type `email` contains `from` or `to`. The combination of origins, categories, types and properties leads to a nested data structure.

When using the Hansken Query Language, both `origin` and `category` are omitted. This means you can find emails using `email.from:'bob@domain.com'`, not via `system.extracted.email.from:'bob@domain.com'`. This also explains why a trace can not contain both `system.extracted.email.from` and `user.extracted.email.from`, as they would point both to the same type and property.

Example:

```
{
  "id": "123", // intrinsic: "id"
  "name": "A trace", // intrinsic: "name"
  "system": {
    "extracted": {
      "email": {
        "from": "bob@domain.com"
      }
    }
  },
  "user": {
    "extracted": {
      "email": {
        "to": "alice@domain.com"
      }
    }
  }
}
```

Toggle between project and default trace model properties: Trace model is shown for the current project

Filter the results on this page:



+	#note	A user defined note.	annotated
+	attachment	An attached block of data.	extracted
+	audio	A media trace that contains audio.	extracted
+	audit	Describes a trace modification	processed
+	browserHistory	Information on a web page or file that has been visited with a web browser.	extracted
+	browserHistoryLog	Log containing the history of visited web pages and files with a web browser.	extracted

Anonymous ▾

Documentation - Trace model

+	file	A file is a block of arbitrary information, or resource for storing information.			extracted
+	fileArchive	A collection of folders and files packed into one file.			extracted
+	folder	A folder is a container within a filesystem, in which groups of files and other folders can be kept and organized.			extracted
+	gps	A GPS location, a possibly timed global position.			extracted
+	identity	An identity is a small contact card that describes identifiers used elsewhere in the trace.			extracted
	image	An image is an artifact, that has a similar appearance to some subject - usually a physical object.			extracted
+	origin	Information about the origin of the trace.			processed
+	picture	A picture is a two-dimensional visual representation.			extracted
-	prediction	A statistical prediction about a trace.			mined
	Name	Description	Cardinality	Collection	Type
	class	The predicted class.			string
	confidence	The prediction confidence, expressed as a number between 0.0 and 1.0.			real
	confidences	The prediction confidences, expressed as a numbers between 0.0 and 1.0.		list	real
	count	A count for the given prediction.type.			integer
	embedding	The predicted embedding.			vector
	label	The predicted label.			string
	modelName	Name of the model that produced this prediction.			string
	modelVersion	Version of the model that produced this prediction.			string
	offset	Offset from the start of this prediction.			real (s)
	offsets	Offsets from the start for the predictions, in seconds.		list	real (s)
	region	The prediction region as a polygon [x1, y1, x2, y2, ...] without wrapping back to start.		list	integer (pixel)
	type	The type of the prediction, e.g. a classification, scene text or barcode.			string
	privileged	Whether the trace has privileged access: suspected, confirmed or rejected.			annotated
+	quality	Information about the quality of the value of a property.			processed

Projects

Project images

Search

Singlefile

Browse

Visualization

Code Notebook

Statistics

Preferences

Health

Tasks

Extraction tools

Operations

Documentation

Extraction Plugin SDK

Guide

Python API

REST API

Trace model

Supported Formats

Release notes

Licenses

- Anonymous
- Projects
- Project images
- Search
- Singlefile
- Browse
- Visualization
- Code Notebook
- Statistics
- Preferences
- Health
- Tasks
- Extraction tools
- Operations
- Documentation
- Release notes
- Licenses

Project images

Search 2 [Create Image](#) [Link image](#) [Share keys](#)

Create filter

Description	Key	Created Date	Conversion status	Data locations	Raw size	Stored size	Type	Version	Upload NFI	Uploaded
demo-data.zip	Extract	2023-03-01	incomplete	1:local	100.17 MiB	91.47 MiB	NFI	2.4		✓
dfrws pics										✓

Extract

Extracting traces from an image will delete all previous extraction results for this image.

Type of extraction: Defaults

Tool timeout: Defaults

Map-Reduce timeout: Defaults

Map-Reduce memory: Defaults

Image source: Defaults

[Extract](#) [Cancel](#)

Presets: Presets...

Tools to apply:

- document/doc Configure
- document/pdf/revisions Configure
- imaging/barcode/fire Configure
- imaging/classification/caffe-googlene Configure
- imaging/classification/fire Configure
- imaging/classification/firevideo Configure
- imaging/face/fire Configure
- imaging/scenetext/fire Configure
- internal/mark-privileged Configure
- internal/mark-privileged-terms Configure

- Home
- Case alerts
- Search
- Timeline (Beta)
- Tags and Notes
- Overview
- Multimedia
- Audio
- Photo camera
- Pictures
- Video
- Smart classifications
- File explorer
- Locations
- Accounts
- Communications
- Browser artifacts
- Reports
- Entities
- System


























Smart classifications

- Faces
- Barcodes
- Identifications
- Drugs
- Military
- CT related
- Heavy weapons
- Firearms
 - Assault gun
 - Assault rifle
 - Firearm
 - Revolver
 - Rifle
 - Six gun
 - Six shooter
- Other

Search more


Firearms

100 of 1.628 Traces

 gun.jpeg 11.22 KB	 gun.jpeg 11.22 KB	 13594455...829.jpeg 141.97 KB	 c7519e62...38dfb1f6 8.19 KB	 4581154D...895F4E5D 169.9 KB
 telegram...53-0-0-0 10.75 KB	 f_00023f 27.45 KB	 173706c7...63d.jpeg 251.12 KB	 420af4af...5553ae.0 6.08 KB	 b460f321...0e9335.0 29.52 KB
 table.th...dex:7216 2.74 KB	 table.th...ex:11971 2.74 KB	 table.th...dex:7493 2.74 KB	 2-033_13...-10.jpeg 11.06 KB	 Standard...ed_9.jpeg 11.06 KB
 2-033_13...-10.jpeg 11.06 KB	 2fe47150...181088f1 47.33 KB	 table.th...dex:4577 3.81 KB	 8ddc5c44...213e1e63 71.06 KB	 table.th...ndex:983 2.8 KB
				

Trace details

PREVIEW HEX VIEWER



TRACE INFORMATION NOTES (0) ENTITIES

EXPORT ADDED BY ANALYZING TOOLS

SMART CLASSIFICATIONS

Trace information

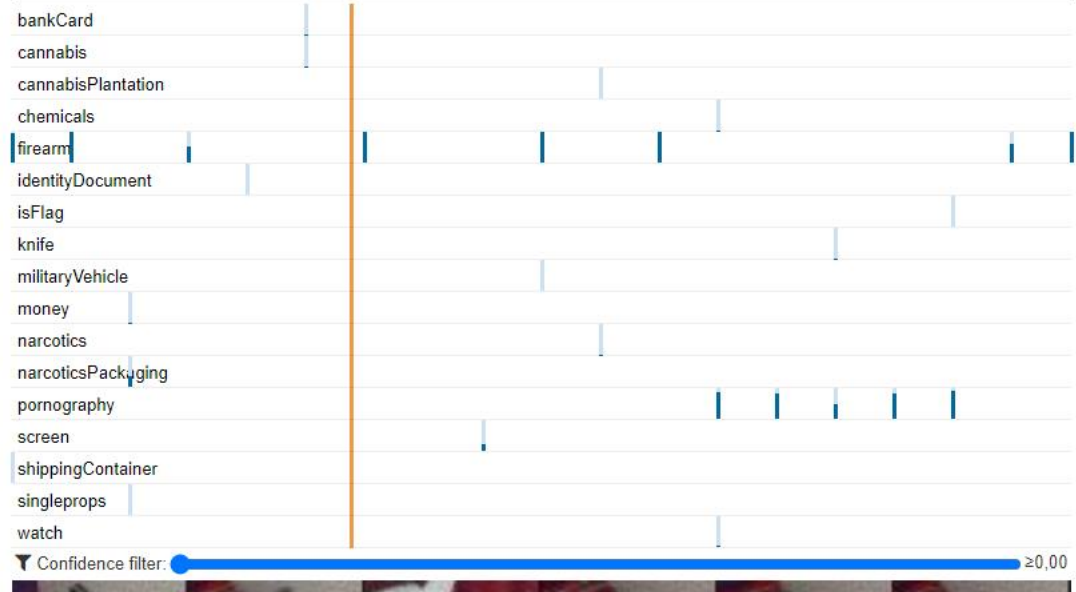
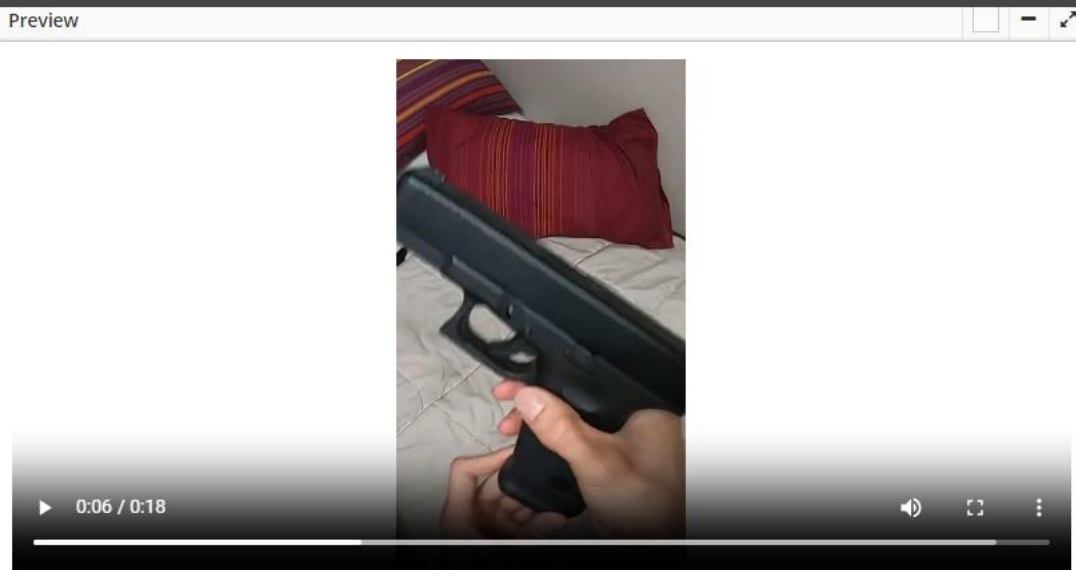
Name	gun.jpeg
Size	11486 bytes (11.22 KB)
Created	20-07-2022 07:49
MD5	e561c5bde96e1f76ae873790ae14bd06
Path	MacBook EN (logical) (3496) / AD1 / Users / chapp / Documents / How to get away with murder / gun.jpeg
Camera	
Taken	
Size	300 x 168 px

extracted

mined

processed

- Anonymous
- Projects
- Project images
- Search
- Singlefile
- Browse
- Visualization
- Code Notebook
- Statistics
- Preferences
- Health
- Tasks
- Extraction tools
- Operations
- Documentation
- Release notes
- Licenses



- Home
- Case alerts
- Search
- Timeline (Beta)
- Tags and Notes
- Overview
- Multimedia
- File explorer
- Locations
- Accounts
- Communications
- Browser artifacts
- Reports
- Entities**
 - Overview
 - Addresses
 - Advertising IDs
 - BitLocker
 - Certificates
 - Document author
 - Imei
 - Phones
 - System

- Entity type
- Search
- All
 - Url (8937372)
 - Emailaddress (6328816)
 - Phonenumber (1717515)
 - Macaddress (1312468)
 - Whatsapp (1280110)
 - Bsn (704294)
 - Ipv4address (508867)
 - Creditcard (433441)
 - Ipv6address (161369)
 - Cve (105779)
 - Imei/phase2 (59067)
 - Imsi (54820)
 - Windows/registrykey (47046)
 - Coinaddress/ethereum (13587)
 - Imei/phase1 (12059)
 - Licenseplate/dutch (1732)**
 - Seedphrase/bip39 (974)
 - Coinaddress/bitcoin (493)
 - Seedphrase/electrum (165)
 - Iban (116)
 - Coinaddress/litecoin (91)
 - Coinaddress/monero (35)
 - Coinaddress/bitcoingold (8)
 - Coinaddress/dogecoin (5)
 - Coinaddress/ripple (4)

Entity overview by trace
















SHOW ALL ENTITIES

Search

LOCAL FILTERS


LIST VIEW

110 of 110 Traces

 Unknown	 Unknown	 Unknown
 Unknown	 Unknown	 Unknown
 Unknown	 Unknown	 Unknown
 Unknown	 Unknown	 Unknown
 Unknown	 Unknown	 Unknown

Trace details

PREVIEW HEX VIEWER



TRACE INFORMATION NOTES (0) ENTITIES EXPORT ADDED BY ANALYZING TOOLS

SMART CLASSIFICATIONS

Trace information

- Name: 3EAF012108BF3206477BD47FD531E17A4A293937
- Size: 354676 bytes (346.36 KB)
- Created: Unknown
- MD5: dafa2d1dab94975903b19b555bd98c28
- Path: [Samsung_Galaxy_J7_\(8411\) / LOCAL / Samsung_GSM_SM-J730F_Galaxy_J7_Pro_2022-08-01_Report.ufdr / File Root / Samsung_GSM_SM-J730F_Galaxy_J7_Pro.zip / data / data / org.mozilla.firefox / cache / x22ewcho_default / cache2 / entries / 3EAF012108BF3206477BD47FD531E17A4A293937](#)

Camera

- Taken
- Size: 1920 x 1279 px

- extracted
- mined
- processed



Case investigators

know all about their case

are not digital experts

can use a computer

understand digital traces

might jump into conclusions

don't search per evidence item
or any other structure

find potential evidence



Digital experts

typically support cases

often do repetitive work

can use expert tools

can dig much deeper

understand digital evidence

data structures direct investigations

report actual evidence



training, training,
training, ...

Case investigators

know all about their case

are not digital experts

can use a computer

understand digital traces

might jump into conclusions

don't search per evidence item
or any other structure

find potential evidence

service all people involved, they all have different needs and wishes

DFaaS NEVER replaces
digital experts

Digital experts

typically support cases

often do repetitive work

can use expert tools

can dig much deeper

understand digital evidence

data structures direct investigations

report actual evidence





Hansken's main characteristics

Governmental

- developed and maintained by Netherlands Forensic Institute ✓
- Steering committee by Dutch governmental partners ✓

Transparent

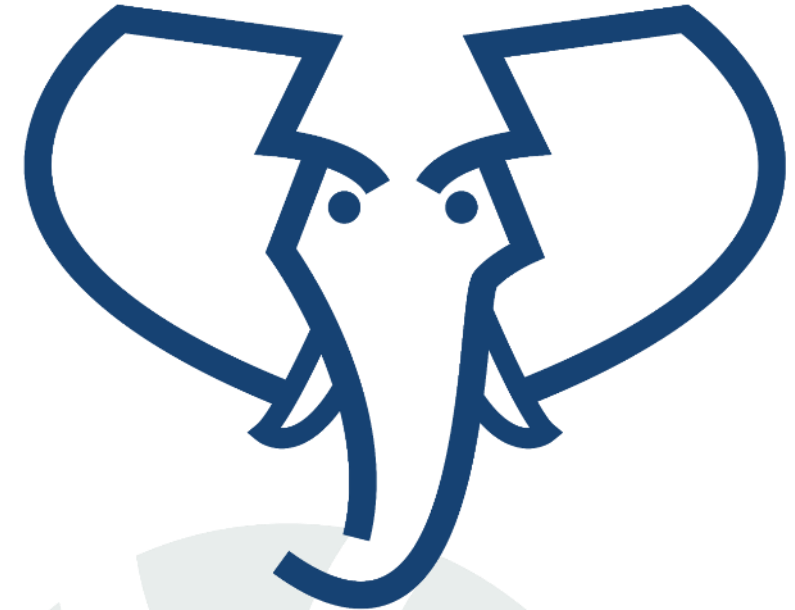
- all evidence is traceable (chain of evidence) ✓
- advanced logging of embedded data processors ✓

Scalable

- > 2000 cases, > 100 concurrent cases ✓
- cases with > 1,000 seized devices ✓
- cases with > 700 terabytes of data ✓
- cases with > 1,000,000,000 traces ✓

Legal

- all handling of traces is auditable (chain of custody) ✓
- support for handling privileged communication ✓
- profound judicial review in Dutch courts ✓



Hansken

The open digital forensic platform
investigate - innovate - share



Building the Hansken Community in the Netherlands....



Netherlands Forensic Institute
Ministry of Justice and Security



NETHERLANDS
PUBLIC PROSECUTION SERVICE



Ministry of the Interior and
Kingdom Relations



Netherlands Food and Consumer
Product Safety Authority
*Ministry of Agriculture,
Nature and Food Quality*



FIOD
Belastingdienst



Royal Netherlands Marechaussee



Netherlands Labour Authority
Ministry of Social Affairs and Employment



Human Environment and Transport
Inspectorate
*Ministry of Infrastructure
and Water Management*



POLITIEACADEMIE

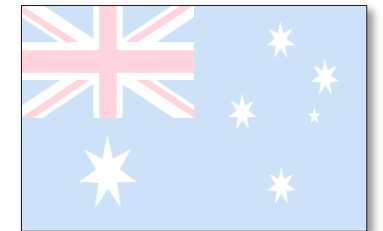
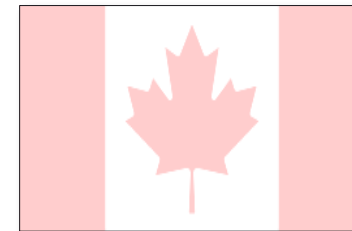
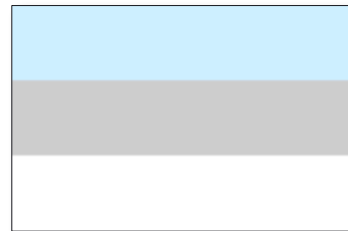
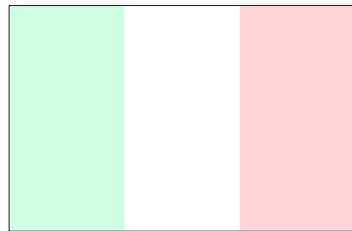
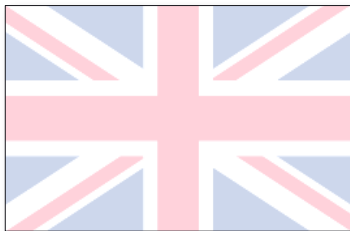
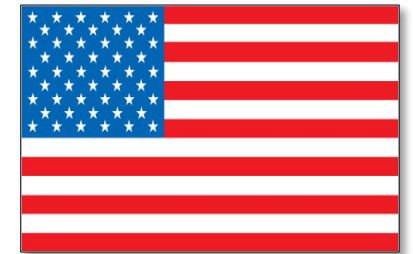
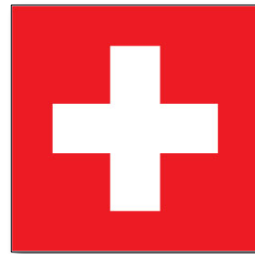
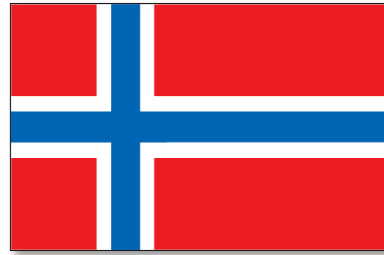
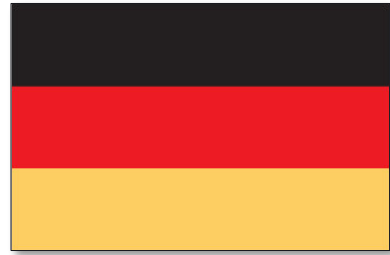


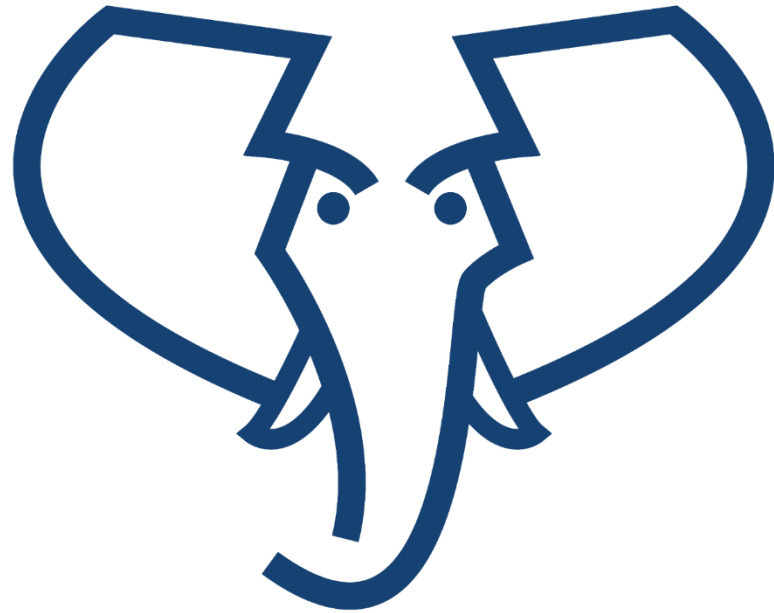
Justitiële ICT Organisatie
Ministerie van Justitie en Veiligheid





Building the international Hansken Community....





Hansken

The open digital forensic platform

investigate - innovate - share

Hansken is the open digital forensic platform

- Extensible by all users
- Parts can be replaced

Investigate (LEA only)

- use the platform for crime case investigations
- Use the platform for intelligence investigations
- efficient mobilization: case investigators, digital experts, analysts, data scientists, advocacy, judges

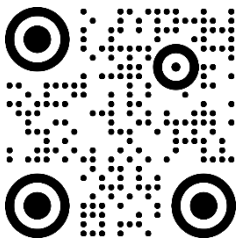
Innovate

- R&D with focus on extending DF knowledge
- continuous development of the platform itself

Share

- Hansken is a knowledge “hub”
- share DF methods and procedures
- share DF tools and technology

extended version is available at <https://www.hansken.nl/hansken-community/vision>



Launch of new Hansken facility: lawyers can now view digital traces from their own workplace

News item | 20-03-2023 | 12:33

Together with the police and the Netherlands Public Prosecution Service, the Netherlands Forensic Institute (NFI) has developed a method that enables lawyers to view crypto communications from their own offices, like for example the Encrochat data. This is accomplished by using Hansken, a digital forensic search engine. Martijn Egberts, the national public prosecutor for digital detection, states: “This new



Thoughts on using AI for forensic purposes

Hansken is:

- used for investigations use, but
- designed for evidentiary use

Evidentiary use is more strict:

- Accurate
- Repeatable
- Reproducible

So algorithms must be:

- Explainable
- Validated
- Deterministic
- No external data

Artificial Intelligence:

- Use external data: Training sets
- Use external data: Cause bias
- Lacks explainability

Use AI for investigative purposes, with:

- Disclaimer
- Education

by the way:

- Not all currently used algorithms are good
- Data under investigation can result from AI (e.g., consumer product AI)

<https://blog.ampedsoftware.com/2021/10/05/can-ai-be-used-for-forensics-and-investigations/>



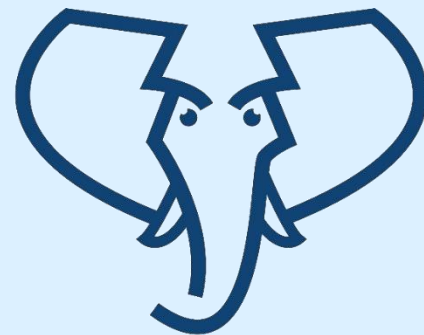
To be continued...

Harm van Beek PhD
harm.van.beek@nfi.nl



Webinar:
Algorithms in Forensic Science:
Challenges, Considerations,
and a Path Forward

<https://forensicstats.org/>



Hansken

The open digital forensic platform
investigate - innovate - share

DFaaS papers

<https://hansken.org/>



international standard for
automated combination,
validation, and analysis of
cyber-investigation information

<https://caseontology.org/>