INSPECTr Project

Intelligence Network & Secure Platform for Evidence Correlation and Transfer

Quarterly Newsletter: Eighth Edition

Edition: February 2023

Intelligence Network & Secure Platform for Evidence Correlation and Transfer

Principal Objectives Brief Summary

To develop a shared intelligence platform and a novel process for gathering, analysing, prioritising, and presenting key data to help in the prediction, detection, and management of crime in support of multiple agencies at local, national, and international level. This data will originate from the outputs of free and commercial digital forensic tools complemented by online resource gathering. The final developed platform will be freely available to all Law Enforcement Agencies (LEAs).

INSPECTr Newsletter Eighth Edition

Welcome to the INSPECTr project newsletter, a guide to our latest work and news. In this our eighth and final edition we will provide our last quarter activities including a Blog on the INSPECTr Project's final dissemination event, **A Series of CEPOL Webinars to Demonstrate the INSPECTr Platform**, an update on our final **Living Lab Experimentation Phase (Phase 6)** and news about the INSPECTr project's other dissemination events and activities.

BLOG

A Series of CEPOL Webinars to Demonstrate the INSPECTr Platform

In February 2023 the INSPECTr project presented a series of lunchtime webinars, courtesy of CEPOL, with the target audience being Law Enforcement Officers, Judicial Authorities, and EU Public Security Entities fighting cybercrime.

The main goal of the INSPECTr project is to create a proof-of-concept platform, but future development will aim to improve the technology towards operational use so that it will be adopted by European LEAs. The platform can be used for a wide range of LEA activities, such as digital forensics and open source intelligence gathering. However, it also addresses major issues that LEAs experience, such as big data management and collaboration with other jurisdictions. The aim of this webinar series was to inform the target audience on how the INSPECTr platform can be accessed, installed, and configured, how to use INSPECTr gadgets for the acquisition and processing of digital evidence and intelligence sources, and to demonstrate the platform's case management system and analytics services. Comprehensive practical demonstrations were presented throughout the week-long webinar series. Access to the recordings of the webinar series is available to those registered on the LEEd platform, CEPOL's online education and training platform.

1. Project Overview, Platform Setup and Usage

This webinar described the platform and outlined how the needs of LEAs can be continually addressed at low cost.

INSPECTr provides:

- Fusion of outputs for commercial tools.
- Integrated tools for digital forensics and intelligence gathering.
- Assistance via AI/ML driven cognitive approaches.
- Proactive policing techniques for detecting and forecasting crime.
- Extra-jurisdictional collaboration: correlation and discovery of evidence.
- Full chain of custody and stack using blockchain ledger and services.

The project has been focussing on major LEA investigative issues which typically often come from having to deal with the huge volumes of data that places huge pressure on LEAs. Digital forensics units have to process data across a broad range of crimes, not just those concerning cybercrime. There is also a broad range of tools being utilised within forensic units for various purposes which often produce different outputs. This makes performing homogenised analytics on the data very difficult. There are also budgetary restrictions to be considered regarding purchasing, training, and licenses, all of which can create backlogs and bottlenecks in processing, and legal, technical, and bureaucratic obstacles in the way of LEAs wishing to work cooperatively together across borders. To ensure that the INSPECTr platform has developed to meet the needs of law enforcement, the INSPECTr law enforcement partners have been involved in the research and design and development

of the platform from the outset and were key in the co-creation of three mocked use cases, and for providing invaluable input in the 6 phases of platform experimentation, testing and feedback cycles that have informed development.

The INSPECTr project has been developed with linkage to other projects:

- **FREETOOL project** – The Freetool project has a suite of tools that have been developed by law enforcement and are free for law enforcement only.
- **CASE Ontology -** Cyber-investigation Analysis Standard Expression (CASE) is an ontology that is used to standardise the output of INSPECTr tools. CASE is essentially JSON linked data that provides a lot of things LEAs need like chain of custody and chain of evidence. CASE allows all tools to output the same language which can then be fed into the INSPECTr analytics system. While integrated free tools are required to produce CASE output, the outputs of existing commercial tools like UFED and AXIOM are also parsed by INSPECTr, to be CASE compatible, resulting in a full LEA toolset that is potentially fully compliant.
- **SIREN -** The SIREN platform is the INSPECTr analytics platform allowing for data to be presented in a human readable format.
- **eCODEX** is used to connect LEA nodes and is key to supporting the discovery of evidence between platforms.

System Architecture

LEA Nodes - Each LEA node is a server that can be purchased at low cost, and each has the INSPECTr platform deployed to it. This includes all of the tools, services, and storage layers, which are free and without requiring any commercial licence.

INSPECTr Gadgets (analysers)

INSPECTr's docker orchestrator is based on Cortex, an existing platform for running security incident response tools. We have developed some modifications to this platform and have developed libraries to provide more functionality. We have a blockchain for immutable logging and we also have data standardisation for analysis, chain of custody and chain of evidence using the CASE ontology. For storage we have Elastic storage for JSON information, and we have Hadoop for storage of binary files, and we have neo4J which allows us to examine the relationships between information. The deployment strategy used by INSPECTr makes it easy to add new tools, fix bugs and respond to feature requests. Investigator access to the dockerised tools can be controlled using Cortex, in order to comply with local policies and legal requirements. This session concluded with a demonstration of how to set up an investigation in the case management system (CMS) and create data within datasets that can start the investigation. Once an artifact was added, the gadgets available to that specific datatype were displayed.

2. Featured Tools and Basic Data Visualisation

The second webinar presented how INSPECTr provides a range of free tools and commercial tool parsers in a single homogenised environment.

OSINT Tools - The platform provides a selection of useful tools for gathering information from online sources. Many of these were demonstrated during this session, which was based around a network intrusion investigation. In addition to processing artifacts in the CMS, the reports produced from the gadgets were showcased, as was the ability to import new artifacts from the reports, for further processing. This provided a simple example of how the CMS could be used to conduct a full investigation into a suspicious IP address using the OSINT gathering tools.

Triage and Digital Forensics Integrated with INSPECTr System - INSPECTr Blockhasher aims to address challenges in LEA units and offer the improvements of identification of target devices (digital triage/preliminary analysis), which can then be prioritised for full forensic analysis. Improvements have been made in the speed of how a device can be prioritised for further examination by providing fast, simple, data classification at block level thus avoiding delays and bottlenecks of full file extraction and hash matching to known files of interest. Data minimisation and storage requirements have also been a focus of development, as only target files will be extracted and validated.

Deepthought was initially developed as part of the FREETOOL project and was selected for integration with the INSPECTr platform. This has since been integrated into each INSPECTr node as a Gadget. Deepthought reads an entire forensic disk, obtains a listing of all the files and then recovers them. The files are then stored onto the actual internal storage and makes them available in the CMS for further processing. It also does further analysis on all the files obtained by providing metadata such as location data, modified times, etc. The processing of large forensic images takes time, but the advantage here is that the process will run in the background, uninterrupted by other systems.

Commercial Tool Report Parsing - There are a number of commercial tool reports currently being processed by Gadgets; currently Axiom, UFED, Oxygen and XAMN. As these tools produce reports in various formats, they must be parsed by INSPECTr gadgets in order to homogenise their data. This will allow the artifacts produced to be subjected to further processing while also supporting the analysis and correlation of evidential data . A lot of new entities were added to the CASE ontology in order to allow for this, by feeding INSPECTr requirements back into the CASE community.

Report Parsing - As well as commercial forensic reports the platform lends itself to adding other types of parsers due to the nature of gadgets. If LEAs have a script or unusual file type, it is foreseeable that a gadget can be added to deal with a unique file type, thus “gadgetising” any process for parsing or processing of files. For example, automatic number plate recognition logs and credit card statements can be used in the system, due to additional parsers being developed for these types of file formats, at the requests of our LEA partners.

3 Data Standardisation, Chain of Evidence/Custody and Analytics

The third webinar presented how the INSPECTr platform provides a range of free tools and commercial tool parsers in a single homogenised environment enabling all tools in the platform to report using the CASE (JSON-LD) format.

With standardisation being a key concept within the project, from the outset the INSPECTr project opted for the open-source CASE/UCO ontology to serve as a standard for interchange, interoperability, and analysis of investigative information. To perform digital investigations effectively, there is a pressing need to harmonise how information relevant to cyber-investigations is represented and exchanged. CASE provides a structured specification for representing information that are analysed and exchanged during investigations involving digital evidence. Moreover, CASE enables the merge of information from different data sources and forensic tool outputs to allow more comprehensive and cohesive analysis. Standardising how cyber-information is represented addresses the current problem of investigators when they receive relevant information from different sources in a variety of formats. Finally, the standard maintains provenance at all phases of cyber-investigation life-cycles, including chain of custody and chain of evidence.

An investigation generally involves many different tools and data sources, effectively creating separate storerooms of information. Manually pulling together information from these various data

sources and tools is time-consuming, and error prone. Tools that support CASE can extract and ingest data, along with their context, in a standard format that can be automatically combined into a unified collection to strengthen correlation and analysis. This offers new opportunities for searching, contextual analysis, pattern recognition, machine learning, and visualisation. Furthermore, organisations involved in joint investigations can share information using CASE.

Despite the challenges, standardisation is key to LEAs and other potential stakeholders in the investigative digital age scenario for integrating and validating the tools LEA are using; providing unified, even federated, data analytics; evidential integrity, secure and reliable exchange; providing interoperability with other projects/platforms; encouraging vendor compliance. If LEAs are currently using offerings from commercial vendors, we hope that soon they too will support CASE as this will be better for everyone, enabling the ability to perform cross-checks more easily across multiple different tools and formats and removing the need for users to be locked into specific tools.

SIREN Platform Overview - The SIREN platform is a component of the overall INSPECTr investigation platform used for data exploration and visualisation. It is a modern investigative intelligence platform that will read the JSON-LD produced by CASE and present it in a human readable format. SIREN uses a data schema (ontology) to coherently coordinate business intelligence style dashboards with best-in-class full text search, knowledge graph link exploration, domain specific visualisations and more. A data model is created by a user specifying multiple cross-index relations. When a query is sent to Elasticsearch requiring data from multiple indices, Siren Federate uses the relations in the data model to construct the appropriate join queries. Based on this data model, dashboards and visualisations can be created to show and connect all distributed data created from jobs created by INSPECTr gadgets, thus allowing the CASE ontology to be represented in a really understandable way. SIREN investigate is the front end and this whole platform sits within the INSPECTr node. This is where you can build relations between datasets, build and view dashboards for visualisations, and where you can analyse and explore data. It visually represents data that has been processed via the CMS and when you start off the gadgets the output of those can have all the data represented in a much richer visual environment. Additionally, it not only represents the data that has been directly processed, but because we operate with the output of many vendors such as UFED/ AXIOM, it means that INSPECTr can represent those in a unified way and there is a centralised analytics platform where all of these can be viewed together, rather than having to use different pieces of software. Every device is displayed together, and we can cross-reference devices and do searches on groups of devices. SIREN Federate can also connect to existing data sources and pull data, not just the data pushed through the CMS, therefore providing linkage to a law enforcement agency's own pre-existing data. A demonstration followed of SIREN's formal representation of the entities and relationships that exist in a domain, creation of connections between artifacts created from gadgets that were run in the CMS of INSPECTr, and how it therefore provides a ready-to-go solution for law enforcement forensic analysis.

4. Integrated AI/ML Tools

This webinar covered INSPECTr Visual (Image Processing), INSPECTr AI Framework, AI Toolbox and SIREN AI.

There are several image processing gadgets available in the CMS; for example Nudedetector, Nudeclassifier, Facial Recognition, Deepface, Childclassifier and Optical Character Recognition (OCR). While the goal of all INSPECTr tools is to generate CASE, there isn't currently a full representation for machine learning in the CASE ontology. It is a difficult format to deal with as there are so many different types of machine learning models, and so many disparate outputs. In order to resolve this issue, we have used the rules of the CASE ontology to generate and create our own structures to

generate this information. The models themselves are stored locally, and held in the local docker registry to avoid the overheads of fetching large volumes of data when needed.

Image Processing Architecture - Cortex is at the back end of the CMS, where all the analyses happen when we execute a gadget. It pulls all of its docker images from the local registry on demand. We also use FastDeploy and Machinery for this framework, which is essentially another type of docker orchestrator specifically for interacting with ML models. SIREN can then be used for providing the 'big picture' of all the returned information contained in all the analyses, presenting the information in a more understandable and usable manner, and so making analysis of all the information easier. Processing of machine learning can be resource intensive, sometimes requiring a lot of GPU/CPU resources, and so we use the INSPECTr visual stack to queue all the data for processing, to avoid all the files having to run at the same time. However, this has not negatively impacted the processing time. Any model being developed and trained requires further feedback and development for it to remain accurate.

There is an important ethical side to Machine Learning to be considered. For example, it must be ensured that the data being used for training models has received ethical scrutiny, to avoid bias or prejudice. It must also be ensured that datasets were gained with correct permissions, and that the information covers a wide base of the source materials.

AI Toolbox Architecture - The AI toolbox is a set of specific purpose tools that have direct access to functionalities used internally and are for specific and occasional use. LEAs do not need to use them automatically, but they can be used if required during an investigation. This is a micro services-based approach with a simplified user interface for easy access, and REST Microservices can be used by existing systems. This platform is a very large sub-set of the French Gendarmerie's that won the Europol most innovative project of 2022 award. While designed for use by individual law enforcement officers, the system can also be used to provide services, such as language translation, to legacy systems. There is a micro-service where you send information and it gives you the answer, thereby increasing the effectiveness of such LEA systems.

The collection of AI Enrichment Platform Tools was introduced and demonstrated, and practical examples provided of how each of these tools can be applied.

- **Multilingual Neural Machine Translation** – a service that can be completely offline, installed in your own internal network and with the possibility of adding and retraining models.
- **OCR** – this follows the same model as the INSPECTr visual. It goes a bit beyond the OCR as it extracts all the information of the file, e.g. in the case of a PDF, it extracts the entire text.
- **Entities and Semantic Links Detection** – you can request entity recognition annotation and add new classes, using these to train your model, and then create the available links and relations for further analysis.
- **Speech to Text** – the tool performs automatic speech recognition, detects and translates the language, and has a high accuracy rate, even when the recording sound quality is poor.
- **Stylometry** – identifying a user's way/style of interacting, who may be active across different platforms, based upon the way they write.
- **Knowledge Discovery** – a way to analyse relations existing between data. It is an exploratory tool to help better understand the data that you have in a simple and fast way so that you can pursue the most relevant parts.
- **Other models** are in process of being added, such as **object detection** and **scene detection**.

Using the INSPECTr framework, it is relatively easy to add additional own models, if LEA require further services. Many of the tools demonstrated here are also available via Europol's Innovation Lab.

5. Evidence Discovery and Exchange, Other Features and Future Exploitation

Evidence Discovery and Exchange: We have so far looked at how digital forensic tools can extract media from evidential material and how machine learning and AI tools can provide cognitive processing of the media, which in turn enables the investigator to query the metadata obtained. In the fifth webinar we look at the transfer-messaging part of the INSPECTr system, that enables one LEA to be able to highlight something to other LEAs on the network of INSPECTr nodes, thereby providing the ability to discover information between INSPECTr nodes. The CASE standard common language will be used to structure and support the automated normalisation, combination, correlation, and validation of information. Additional constraints and privacy considerations highlighted by law enforcement feedback would need to be integral in the Publish/Subscribe (pub/sub) engine's configuration.

Pub/Sub Request Engine – Sending a Request Using Hashes - In this "mocked" example the image classifiers have examined images recovered from a USB device and their combined outputs (nude detection and age detection) have identified possible Child Sexual Abuse Material (CSAM). At this point others in the INSPECTr network can be asked if they too have seen this file as part of their own investigations. To maintain data privacy and reduce exposure to such media, a hash of the file can be used in a query to other nodes. The Pub/Sub query is created to ask, "have you seen this hash anywhere across your investigations?". If one of the members has encountered the hash, then they have also seen the same image in an investigation.

A demonstration followed of the practical steps of how a Pub/Sub request could be created and sent, how these can be managed via the Information Request Management Engine (IRME), how to select where the request is sent, i.e., to which LEAs, how to include messages, how to add time-outs in which to receive responses, and also how to amend the status of the query as it is progressing through the system. Importantly, it was also demonstrated how a strict approvals process from admins/superiors is required throughout the entire process.

Rules Engine: There is a Traffic Light Protocol (TLP) system in-built that takes place for each incoming message. In the background there is a rule engine to check if it is OK to exchange information on these types of crimes, with the various LEAs in the network. Then, depending on the rules in the rules engine, which depend on what bilateral agreements may be in in place between different LEAs and countries, we will get a response on whether the exchange of information is allowed or not. If it is not allowed by the rules engine, then it is not possible to proceed. The final step is for an administrative approval to take place, before routing the request to other nodes.

eCodex: A secure and encrypted routing mechanism, and the main component of the Pub/Sub.

- The Pub/Sub component exchanges data using **eCodex**, a secure, encrypted, and automated routing mechanism.
    - Uses **eDelivery** to exchange documents and data among different systems via a common protocol.
- **eCodex** provides the technical infrastructure to connect different legal systems and ensure fast access to justice past borders.
- Solution deployed as a "black-box" on Windows.
- However a Linux version has recently been released, which will be used in the future.

Important to Note: **Requests can only come from an investigation already in the case management system**. Due to the way we have implemented **CASE**, **Blockchain** and **INSPECTr Gadgets**, we have data being created in an investigation that is fully recorded, in terms of who has placed it there, and who has control over seeing it. Furthermore, a lot of thought has gone into how we allow the investigator to find out information from other countries. For example, the legal agreements that must be observed and the administrative controls in place, are all designed to avoid potential misuse by an investigator.

The Blockchain Structure - The blockchain structure provides an indisputable time series of different data points with each one connecting to the previous one. It is a chain of blocks and when a new block is added, all members on the network receive the data point. Having the previous block's hash, they calculate the new block's hash. This is important, as due to this chain of hashes, tampering cannot go undetected irrespective of how far back the tampered data is.

Whenever new data is added on the platform, a request is sent to the Storage Service to "write". The Storage Service logs this request to its logging system sending it to a central logging platform and also takes a hash of this log sending it to the local Blockchain ledger for storage. When you see a log in the central logging platform you also have the hash of that log. The ledger also sends every Nth hash created to the public Ethereum ledger, thus providing a second layer of guarantee, and making it impossible to alter the log without detection during audit.

By using the connections of all these things, Logging, Blockchain and Pub/Sub, we will be able to provide an indisputable log of activity. For example, imagine a seized device has an email sent from it by a suspect under investigation for terrorism offences. The investigator recovered this email via a given tool at a certain time, the artifact and attachment was analysed by the various other AI tools and, once terror related activities were detected, a pub/sub query was created and directed to LEA members on the network, this request was approved, and an LEA responded to this pub/sub query on the network. All these things are recorded in the storage system, and all backed by the blockchain ledger.

Knowledge Graphs

- The graph representation of information recorded in data that is connected.
- Stored in Graph Databases (neo4j) or RDF/triple stores.
- INSPECTr uses **neo4j** for information stored in graph form and is part of the storage service.

Ontologies - An ontology represents information of a particular subject area using a graph form. It does what a knowledge graph does connecting different pieces of information together on a particular subject. We are using the CASE ontology in the INSPECTr platform, and this ontology connects, in a specific structured standardised graph, the information relating to investigations of digital evidence. It can record investigative actions and data sources thus providing chain of custody allowing us to not only know what was found but also to record information on who initiated the processing, which tool was used to find it, and from which piece of evidence.

SKG – The SKG sits inside the graph database and knowledge graph creation and has a native graph approach for consuming CASE data. Its functionality is interlinked with CASE, via the case_builder libraries developed in the project. Modifications can be made to the graph by implementing proof of concept modifications to the CASE ontology, and this change in the graph is backed by a CASE like file that can be stored in the system and can be reported like any other CASE file; i.e., logged to the blockchain ledger. The Graph has version control, and you can get the structure/form of the graph as

it was at any given point. When new things are added in neo4j, the storage system is informed, and logs are written into the blockchain.

Pub/Sub: While the Pub/Sub is useful for discovery of data on other nodes, the entire process will be built around existing processes for conducting joint investigations between law enforcement. We do not intend to replace Mutual Legal Assistance Treaty (MLAT) or European Investigation Orders (EIO) but we want to speed up the discovery of information across borders and jurisdictions, so an LEA can decide who they want to pursue an MLAT/EIO with. In the future, the **knowledge graph** might lend itself to not just putting together two law enforcement investigations bureaucratically or legally onto one investigation, but actually merge data together in a manner that would enhance both sides of the investigation.

Future Intentions: All that has been presented has been to inform the audience about what we have been working on. The overall aim and intention is for all of the above to be seamless and some more work is still required for this to be a mature approach that is useful for LEAs. CASE has been evolving while we have been engaged in the project and it has been difficult to provide a definitive solution while the ontology is going through major updates, many of which have been triggered by our work in INSPECTr. Graph is very, very powerful and so the idea is that we can query a graph rather than a traditional database that would take time and processing, and with using graph we can ask queries and have those answered much more natively. All of the above has been extremely interesting to develop and present in terms of what is possible in the system, in the way we are using the data and what may be possible for INSPECTr to do in the future.

ELSI Ethical Oversight of the INSPECTr Project

All of the above activity and development of the project has been controlled by active ethical governance of the project, the ongoing sensitisation of the consortium to make them aware of important issues, and dealing with challenges as they are presented through collaboration to seek solutions.

Why Ethical, Legal and Society aspects are important:

- Part of **responsible research** and innovation, especially in Horizon 2020 projects and working with LEAs.
- **Ethics**: Technology has an important place in the world and needs to be monitored closely to ensure it is developed properly.
- **Legal**: Due to the nature of the project and intended use of the platform it is important for us to demonstrate and facilitate legal compliance.
- **Societal aspects**: Many people have worries about technology, so we need to create structures and processes to reassure people, avoid misuse, and make technologies that are societally acceptable.

Active ethical governance has taken the form of ELS Impact Assessments, making recommendations in mitigating risks and seizing opportunities to go beyond the baseline, implementing ethical and privacy considerations in the development process of the technology, looking at the organisational processes of using those technologies, and adapting the entire approach specifically to law enforcement.

Main Challenges and Solutions:

- **Appropriate data protection regime**
    - General Data Privacy Regulations (GDPR) and the Law Enforcement Directive.

    - Our conclusion was that GDPR was the most appropriate data protection regime for this type of research. However, the Law Enforcement Directive (LED) could also be used in certain situations.
- **Ethical and lawful data use, especially LEA data**
    - Many discussions were held on datasets and what types of processing might be appropriate in the circumstances.
- **Exploitation Risks**
    - Detailed discussions were held on potential misuse of INSPECTr-like technologies by nefarious actors, reckless users, or well-intentioned people who are unaware of issues, and how these can change over time.
    - A risk assessment was conducted in order to address those potential issues.
- **Limited awareness by end-uses of the ethical and legal issues posed by AI**
    - Developed ethics and legal guidance to cover both the general issues faced by the LEA end-users and also specific guidance for INSPECTr tools.
- **Data Anonymisation**
    - This can be used for decreasing risk for sensitive data and this was discussed with some of the partners and some of the knowledge coming out of that has contributed to standardisation processes.
- **AI Act**
    - Examining the impact of a forthcoming European Union AI Act on technologies like INSPECTr. This is still in development and in flux, so this has been a challenge, but we have tried to 'future-proof' INSPECTr as far as possible.
- **DPIAs and CP-As**
    - Law Enforcement data that is not connected directly to law enforcement investigations can be very difficult to deal with from a legal and ethical perspective. In order to have the possibility to use real closed case data in the later stages of the project for testing purposes, we completed Data Protection Impact Assessments (DPIAs) with participating law enforcement agencies. These were accompanied by Control-Processor agreements (CP-As), which were a legal mechanism to allow some technical partners access to LEA data in some circumstances to provide technical support if needed.
    - Highly detailed templates were developed, which are appropriate for law enforcement agencies for determining where the closed case files are stored and who owns them, holding discussions with data protection officers and data protection authorities, and recognising and mitigating any possible or potential risks to data suspects.

Privacy and security have been front and foremost of the technological development in the project. We have been working to ensure that we can demonstrate this compliance and show INSPECTr to be a responsible project in the way it uses data for testing. The vast majority of INSPECTr technologies have met all relevant requirements and we also consider that we have made good progress towards the INSPECTr platform utilising only 'trustworthy' and 'responsible' AI.

Future Exploitation Discussion and Closing

The project has been quite a challenge to respect all the ethical and data privacy concerns while so much development has been in process, but we have looked at lots of way we can enhance law enforcement processes and existing technologies, whilst meeting the issues and challenges law enforcement are facing, and also future proofing the project. Discovery of evidence across different jurisdictions, the use of AI tools, and even some OSINT tools may cause concerns for LEAs. It was key

to have the LEAs behind us, being focussed across the entire picture, while we were developing the technology and INSPECTr features and ensuring the correct controls around access to the tools, while ensuring full chain of custody and full chain of evidence on all of the tools available.

We have presented a lot of technologies and tools during the week's presentations. However, there were a lot of things we did not have time to present such as online data preservation, web scraping tools, proactive policing techniques for detecting and forecasting crime based on historical data, and so there is a lot more to come from the INSPECTr project.

We set out to develop something that is free for LEAs in keeping with the coordinator's (UCD Centre for Cybersecurity and Cybercrime Investigation) historical ethos of working on EC funded projects to produce free tools and outputs for LEAs across all of Europe. Our next task is to consider what future funding may be available to ensure we can continue to bring the INSPECTr platform forward from the current proof-of-concept, into operational use. Even some individual components of the platform could be valuable as stand-alone projects, that would benefit LEAs. For this we need a lot of LEA support behind us and need every consideration from every corner of Europe factored into the final development phase of this platform and its technologies.

Final Phase of Living Labs Experimentation – Living Lab 6

Living Lab experimentation has continued to be an integral part of the INSPECTr platform's development throughout the project. The final experimentation phase, Living Lab 6, was run over a five-day period, Monday 30th January 2023 – Friday 3rd February 2023.

Local Installation of LEA Nodes: This was the first Living Lab where INSPECTr LEAs were able to run the experimentation exercises on their own locally installed and networked LEA node. Each LEA node had been shipped to the respective LEA and a training session held to familiarise the local IT and network teams with the features of the INSPECTr platform and its capabilities. This training forms part of the capacity building programme of the project and will be available as a training resource for future LEAs, who wish to utilise the platform.

Context: In previously held Living Labs, LEA developed Use Cases that used mocked data had been used for testing and feedback. For Living Lab 6 a slightly different approach was taken with LEA testers asked to work on set specific tasks to test pre-set scenarios. This was to ensure that all the developed features of the platform were tested, and feedback could be collected on them. Although it had been envisaged earlier in the project that the final Living Lab would use some real case closed data, this was not in fact the case. However, if there was sufficient time before the end of the project it remained a possibility that some scaled down LEA testing of real case closed data could be conducted separately. Any LEA testing real case closed data will have completed all documentation required to ensure full GDPR compliance, if any of this required documentation was incomplete testing of real case closed data would not be permitted.

LL6 Structure: Each day commenced with an introductory training session on what the day would entail, followed by a progression through the set exercises, and would end with a section in which feedback would be provided by the LEAs involved in the testing. The Development Team were online and fully engaged throughout each day to work with the LEAs to support with querying, troubleshooting, and fixes in this live testing environment.

INSPECTr Living Lab 6 – 30th January 2023 – 3rd February 2023
Schedule

Day 1: Triage and Digital Forensic Tools

- Part 1: Create Investigations in the CMS
- Part 2: Data Ingestion and Preparation
- Part 3: Triage and Digital Forensics

Day 2: Artificial Intelligence and Machine Learning Tools

- Part 1: Cortex – Image Processor Gadgets
- Part 2: Exercises with the Tools: AI Toolbox
- Part 3: Exercises with the Tools: NLP Gadgets

Day 3: SIREN Demo

- Part 1: SIREN Overview of INSPECTr Data Model and Dashboards
- Part 2: SIREN Demo
- Part 3-4: Practical Exercise and Solution
- Part 5: Working With SIREN
- Part 6-7: Second Practical Exercise and Solution and Wrap Up

Day 4: Data Discovery and Information Exchange:

To query other agencies on the Pub/Sub network, on whether they have come across the same image as in the INSPECTr platform, by using a hash of it when creating a query.

- Part 1 – Sending a Query Exercise
- Part 2 – Receiving a Query Exercise
- Part 3 – Receiving a Reply

Day 5: Other Features

- Part 1: Blockchain – What is the Blockchain / What is a hash?
- Part 2: SKG – INSPECTr platform is ontology based. An ontology represents information in a subject area using fully structured and related concepts in some graph form.
- Part 3: LEA Databases – SIREN - Importing from an external datasource.
- Part 4: OSINT Tools – A practical OSINT exercise was carried out where LEAs were asked to firstly enable their gadgets in Cortex, and then asked to create an investigation, add artefacts, and analyse.
- Part 5: Evidence Visualisation – INSPECTr Project's approach to evidence visualisation using the INSPECTr Electronic Evidence Visualisation Library.
    - A database of templates describing several electronic evidence structure data types in a common language.
    - A binary parser that translates input raw data, by means of the templates.
    - A user interface including a hex-browser (binary visualiser) and a templates editor.
    - Both the tool and the database will be open-source.
    - Useful resource not only for trainees and trainers but also for LEA experts who need:
        - to find evidence in complex data structures and to explain their findings to a non-technical audience
        - to cross-check the results of other forensic tools.

Further Opportunities for INSPECTr Dissemination and Cross-Project Learning and Collaboration

AGOPOL Online Conference November 2022

Title: Can Privacy and Ethics-by-Design by Adapted for Law Enforcement Technologies? Presented by INSPECTr partner Trilateral Research.

Abstract: The impacts that technologies have on us as individuals and on society at large can be significant. It is, therefore, important that technologies are designed and developed in appropriate ways. This is particularly the case with law enforcement technologies due to the exceptional place that law enforcement plays in our societies, especially where data-analysis tools are used to reveal private information about suspects. Two design approaches that can assist in appreciating and mitigating risks raised by law enforcement technologies are Privacy-by-Design and Ethics-by-Design. However, these approaches are primarily focussed on commercial technologies where the end-users are the focus of attention. Yet, with law enforcement technologies, the end user is likely to be a law enforcement officer, such as a detective or crime data analyst, but the focus of attention from Privacy and Ethics-by-Design approaches is the subject of a criminal investigation. How should these approaches be adapted to deal with this change in focus? Another key issue is the lawful ability of law enforcement to uncover private details of individuals present in their investigations, how should Privacy and Ethics-by-Design be implemented in a situation where conventional standards of privacy do not apply, and the standards of what behaviours might be ethical and acceptable are different? This paper seeks to answer these questions and provide an outlook as to how privacy and ethics-by-design approaches can be adapted and applied in the situation of researching and developing data-analysis technologies for law enforcement investigations.

Council of Europe Workshop 7th December 2022 held in UCD, Dublin

This event was hosted in Dublin on the 7th of December and delivered by Ray Genoe and Robbie Dowdall (UCD). During the workshop, 24 participants were introduced to the INSPECTr project and provided with a detailed demonstration of the platform's capabilities. The 11 countries represented included Albania, Armenia, Azerbaijan, Georgia, Kosovo, Moldova, Montenegro, Republic of Srpska (Bosnia and Herzegovina), Serbia, Turkey, and Ukraine. All participants showed great interest in the project and stayed long after the scheduled time to ask follow-up questions about the platform and its capabilities.

ECTEG Experts Meeting 6th – 9th February 2023 in Budapest

40 attendees (mostly LEA experts in digital forensics / cybercrime) attended an event in Budapest that featured the INSPECTr platform. Participants were invited to see the platform in action in one of the workshops and provide feedback on its operational viability.

During 2 sessions, lasting over 3 hours, Ray Genoe (UCD CCI) provided a demo of an OSINT and a Digital Forensics investigation. In addition to the

intelligence gathering tools and digital forensic tools, this provided an opportunity to showcase the new Blockhasher triage tool, the case management system, the analytic dashboards and the AI/ML tools for automated image classification and facial recognition.

Feedback from participants was hugely positive, with participants particularly impressed with:

- The deployment processes for the platform and tools
- The administrative controls for the tools
- The benefits of the platform compared to currently available tools
- The Blockhashing solution to digital forensics triage
- The integrated digital forensic tools and report parsers
- The suite of AI tools available.

CERIS - Spotlight on the fight against Child Sexual Abuse held on 21st February 2023 in Brussels

This workshop presented by DG HOME, aimed to bring together practitioners, local authorities, policymakers, and researchers to discuss latest insights and lessons learned from ongoing EU-funded projects related to the prevention and detection of Child Sexual Abuse, and was attended by the INSPECTr Project Coordinator, Dr Ray Genoe. The aim of this workshop was to strive to ensure that recent research findings carried out can be fully exploited by informing policymakers and other stakeholders, particularly on evidence-based policy responses.

Dr. Ray Genoe, coordinator of the INSPECTr project, was invited to be a speaker at the breakout session on the detection of Child Sexual Abuse Material (CSAM), since the INSPECTr project was developing numerous tools and technologies to support this field of investigation. While many projects, including INSPECTr, are focussed on AI supported technologies for image and video processing, none were addressing the needs of triage/preliminary analysis or the discovery of evidence between jurisdictions for uncovering CSAM distribution networks.

The publish/subscribe process was briefly described but a huge amount of interest was directed to the Blockhasher tool developed by UCD CCI during the project. Additional time was carved out of the agenda to provide a demonstration of the tool to the audience, which included representatives from Europol who are dedicated to combating CSA crimes.

The Blockhasher provides automated triage of digital evidence, so that it can be prioritised for full forensic analysis. This will go a long way to reducing backlogs in digital forensic units, who struggle to process the vast array of devices being seized during such investigations. The automated process is much faster than the current “extract-and-hash” process when dealing with known (bad) files, and can be used by non-technical operators without being exposed to the upsetting nature of the content retrieved. Embracing a data-minimisation principle, the tool simply extracts target files from almost any storage device, thereby also addressing the digital evidence storage issues currently faced by many LEAs.

INSPECTr Final Project General Assembly 24th February 2023 in UCD, Dublin and Online

The project’s final general assembly was held on Friday 24th February 2023 and was a hybrid event of both in-person and on-line attendance. The meeting covered the next key reporting milestones to

be observed by all partners regarding the final financial and technical reporting of the project. There was also a detailed review of the status of all work package tasks and deliverables undertaken. This was followed by a demonstration of the INSPECTr platform highlighting the many features that had been developed to support European Law Enforcement. The meeting concluded with an exploration of the possibilities and opportunities that may exist for further development of the current proof-of-concept INSPECTr platform to one that is a fully operational platform, free-for-LEAs.

Project Activities and Events between Dec 2022 – Feb 2023

- **INSPECTr Monthly Project Meetings**
- **INSPECTr Weekly Technical Meetings**
- **INSPECTr LSG Monthly Meetings**
- **Ethics Work Package Monthly Meetings**
- **EARG Ethics Advisory and Review Group**
- **Living Labs Experimentation Phase 6**
- **Final Project General Assembly**

Conferences and Workshops

AGOPOL Online Conference	Online	November 2022
Council of Europe Workshop	UCD, Dublin	7th December 2022
ECTEG LEA Experts Meeting	Budapest	6th - 9th February 2023
Series of CEPOL Webinars	Online with CEPOL	13th – 17th February 2023
CERIS Workshop (DG Home)	Brussels	21st February 2023
INSPECTr Project Final PGA	UCD Dublin	24th February 2023

Closing

We hope you have enjoyed reading the INSPECTr Project's final newsletter in which we have provided information on our closing activities and an overview of the current iteration of the INSPECTr platform which has been developed throughout the duration of the project.

Although this is our final newsletter, our project website will continue to be available https://inspectr-project.eu/.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833276.