INSPECTr Project
Intelligence Network & Secure Platform for Evidence Correlation and Transfer

Quarterly Newsletter: Sixth Edition

Edition: July 2022

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 Sixth Edition

Welcome to the INSPECTr project newsletter, a guide to our latest work and news. In this, our sixth edition, we will provide updates on our last quarter activities, information about meetings and events attended, our upcoming events, recent dissemination activities and a blog about the most recent Living Lab experimentation undertaken within the project, Living Lab 3.

BLOG

Update on INSPECTr Living Labs Experimentation Phase 3

From the outset of the project the INSPECTr platform was to be built with the active participation of a broad and inclusive pan-European LEA community including the definition of requirements,

feedback and involvement on research and development tasks and the demonstrable use of the outputs.

Living Lab 3 took place in UCD on 21st to 23rd June 2022. This was an in-house meeting where both developers and law enforcement participants were able to meet in person. This proved invaluable from a communications point of view as both groups were able to easily interact with one another during the live testing, allowing both groups to effectively communicate problems and usability issues that needed to be resolved and/or improved in the platform.

An example of this was how the law enforcement participants highlighted usability issues which developers had up until this time been unaware of, but having these brought to their attention, could easily understand. These have since been brought on board and elevated as main priorities by developers to improve or incorporate.

During Living Lab 3 law enforcement participants worked alongside developers as live updates were made to the platform. Developers worked late nights to fix reported bugs, and it was found to be much easier to find effective solutions with law enforcement feedback being available immediately and in person rather than working remotely. The meeting also allowed law enforcement participants a more immediate view of how INSPECTr could fit for them as potential future users of the platform.

Despite fixing many bugs at the live testing event, developers also logged a long list of issues to work on after the meeting. From this work, developers noticed and logged other relevant platform issues to work on in the long term that had possibly not been flagged as requiring attention prior to the live event.

Since Living Lab 3 the CMS (Case Management System) has been integrated on the platform which should resolve a significant amount of the usability issues that were raised. Further work on the graphical representation of data for users continues and is being directed based on law enforcement feedback from the Living Lab.

Publications

Article on An Intelligent Platform for Effective Policing

The INSPECTr consortium was approached to submit an article for EU Researcher Magazine in 2022. EU Research is a dissemination journal focusing on pioneering frontier research. It is published quarterly and distributed throughout 33 countries in Europe to over 50,000 readers. The aim of the journal is to promote research projects to a relevant audience in government and the private sector, as well as academia. In turn this will lead to enquiries of interest, global exposure and dissemination for the projects involved. It is hosted online indefinitely reaching a global web audience. It is read by the key people in national and European governments that control policy and research funding, leading scientific research institutes and major companies across Industries in the private sector.

Read the article here in which Dr Ray Genoe, Director of UCD Centre for Cybersecurity and Cybercrime Investigation, and his colleagues in the INSPECTr project describe how they are developing intuitive, easy to use tools and an intelligent platform that will help law enforcement agencies combat crime more effectively.

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

UNIL Conference – University of Lausanne – Presentation by INSPECTr Consortium member French Gendarmerie

The Center for Artificial Intelligence of the Judiciary Pole of the French Gendarmerie was invited to present its activities to the students and staff of the University of Lausanne. The School of Criminal Sciences sponsored the presentation. The presentation was on the 12th of May 2022, and a great part of the works presented was linked to GN's participation in INSPECTr. There was a 45 minute presentation with 15 to 20 minutes provided for questions. More than just presenting the work being carried out, short demos of some of the tools were provided, for example, the named entity annotator and the stylometry analysis tool, stating that both were improved and fully developed in the context of INSPECTr.

EAFS2022 Stockholm hosted by the National Forensic Centre, a department of the Swedish Policy Authority

EAFS2022 is the biggest European Forensic Science Event. The 2022 event focussed on the transformation of forensic science due to new technical possibilities and developments concerning the fundamentals of forensic science and new categories of crime, and how innovation around forensic approaches is necessary to achieve the results required, facilitated by combining knowledge from different areas of expertise. The INSPECTr project was well represented at EAFS2022 with presentations made by 3 of the INSPECTr Consortium organisations - GN, PHS and CCI/UCD.

1)   Authorship Identification in Web Forums

Presenters/facilitators: Dr Daniel Camara, member of the Center for Data Science of the Gendarmerie, France and Mr Mohamed Belahcen, Data Scientist working in the EU H2020 INSPECTr project, the Gendarmerie, France.

At the EAFS conference, GN presented a method and a framework for authorship identification in web forums. With the growing volumes of data, especially on social media and web forums, the correct identification of authors on Internet Forums is a task that could have great value in cyber investigations. It may help identify different accounts that the same real person controls. In this case, the person may be using multiple avatars to convey one idea or sell a product and other accounts to reinforce the value of their opinion or product. For investigators, this information may help understand the target avatar dynamics and establish investigative tactics.

Stylometry is a technique based on statistics of the text to detect particular writing styles specific to each person and that is hard to disguise fully. Advances in computer science and especially in natural language processing have made the methods much more cost-effective and prone to be automatically applied to large datasets. The techniques may be as simple as counting the average number of words per sentence and the length of these words, or as complex as accessing the semantics of the given texts through sentence embeddings.

Based on stylometry, the suggested model is a combination of both linguistic features and embeddings to identify specific writing styles. The model is based on cosine similarity and yields an average accuracy of up to 99.4% when using 200 texts per author. The framework also shows specific similarities in writing styles whether they are lexical or syntactic.

Mr Mohamed Belahcen, Data Scientist working in the EU H2020 INSPECTr project, the Gendarmerie EAFS2022

2) Natural Language Processing for Social Network Analysis Workshop

In this EAFS workshop, the French Gendarmerie demonstrated the build of a full data pipeline capable of analysing social media posts. Participants learned the theoretical and practical aspects of performing sentiment analysis and detecting hate speech and radical or extremist content on web forums and social media posts. In the workshop, different methods for data exploration and data analysis were discussed. GN presented how to build statistical and deep learning methods for performing sentiment analysis. The same type of method is used in INSPECTr. Participants could see how these methods work and understand different natural language processing methods that could be useful during investigative activities. In a didactic and easy-to-follow way, participants could follow and try the methods during the workshop using a collaborative tool. The objective was to provide the fundamentals of a simple but complete data analysis pipeline in Python. Participants did not require advanced computer programming skills, but knowledge of programming was important to take full advantage of the workshop.

3) Automatic Web Parser

The second paper presented at the EAFS Conference described a tool for automatically creating parsers for web forums and markets. There is no standard format for web pages if we want to extract and organise the information collected for Open Source Intelligence, e.g. name of user, price of a product, type of product, contact information, etc. Therefore a parser is required for this task. A parser is a piece of software capable of extracting the information of a text within a given structure. The problem is that, as the pages are not standard, each website demands the creation of a specific parser for that website. On top of that, the process is subtle, time-consuming, and needs to be done by a computer science expert. GN has worked on a generic and semi-automatic parser generator that, from the repetition of the structures of the website page, can propose elements that may be of interest and are extractable. The user just needs to label the field to give a semantic to it, for example, user_name, and every time the given HTML structure is found over all the collected pages for a site, the information is now tagged user_name. The method needs only one example page and the guidance of a field expert to verify what is important and what is not. After that, all the process of data extraction is automatic.

Dr Daniel Camara, member of the Center for Data Science of the Gendarmerie Presenting for INSPECTr Project EAFS2022

4)   Cross-evidence Crime Analysis Platform for Evidence Correlation and Transfer

Presenter/facilitator: MSc. Yves Vandermeer, lecturer at Police University College (PHS), Norway

In this workshop it was demonstrated how the output of miscellaneous digital forensic software can be homogenised, ingested, and enriched so that crime analysts can discover and validate the hypothesis using visualisation like timeline, comparisons, and cross correlation. Based on a provided mocked crime scenario and related documentation, the workshop started from pre-generated reports from smartphones and computers, part of a mocked crime scenario dataset. Participants were able to compare the original report with the homogenised version and add some non-IT information. Participants then used the interactive visualisation of the data to check two investigation hypotheses with the available evidence.

5)   INSPECTr: Intelligence Network and Secure Platform for Evidence Correlation and Transfer

Presenters/facilitators: Robert Dowdall, UCD Centre for Cybersecurity and Cybercrime Investigation, Dublin, Ireland, Panos Protopapas, Inlecom Group, Brussels, Belgium, Antonis Mygiakis, Inlecom Group, Brussels, Belgium, Ray Genoe, UCD Centre for Cybersecurity and Cybercrime Investigation, Dublin, Ireland

This workshop explained how the EU-funded INSPECTr project will integrate a range of high-tech approaches, including big data analytics, cognitive machine learning and blockchain technologies into

a shared intelligence platform that will improve digital investigations and forensic capabilities, and reduce the complexity and cost of cross-border collaboration.

The platform has been designed through extensive collaboration with the law enforcement community, will incorporate privacy and ethics by design principles, and will take into account relevant national and international legislation. After the project, the platform will be freely available to the law enforcement community and adoption will be enhanced through training courses, webinars, and workshops. Exploitation of the project deliverables will also be freely available to LEAs to further improve the platform beyond the scope of the project, through additional research and development activities.

This workshop also presented the living lab approach adopted by the project and discussed the social, ethical, legal and technical challenges faced by the consortium. The platform's capabilities were also demonstrated. It was explained how the outputs of commercial forensic tools will be combined with the outputs of integrated "free" tools for analysis and correlation. In addition to showcasing various AI approaches to cybercrime investigation, including cross-case correlation, the workshop illustrated the INSPECTr project's approach to evidence discovery and exchange with other jurisdictions.

CEPOL Research and Science Conference 2022, Vilnius. 'Preparing Law Enforcement for the Digital Age'

The CEPOL Conference held in Vilnius in May 2022 invited empirical findings and analytic reflexion on familiar and new digital forms of criminal behaviour as well as on the tools, methods, and strategies to counter them. The promise and challenge of AI has been a topic of great interest, but not less as contributions discussing the potential of digitalisation for training, education and practice of policing and other law enforcement activities. The INSPECTr project was well represented at the CEPOL Conference with 3 presentations by 2 of the INSPECTr Consortium organisations, CNR and CCI.

1) Developing of a Judicial Cases Cross-Check System for Case Searching and Correlation Using a Standard for the Evidence

Presenter/facilitator: Fabrizio Turchi, the technological director at Institute of Legal Informatics and Judicial Systems of the National Research Council of Italy (CNR-IGSG).

In a recent EU publication, a report commissioned by the European Union related to the Cross-border Digital Criminal Justice environment, a set of specific business needs have been identified. Some of the most relevant ones have been:

- The interoperability across different systems needs to be ensured.
- The stakeholders need to easily manage the data and ensure its quality, allowing them to properly make use of it (e.g. use the data as evidence in a given case).
- The stakeholders investigating a given case should be able to identify links between cross-border cases. Therefore, solutions are needed to allow the stakeholder to search and find relevant information they need for the case they are handling.

The study presents a set of solutions to address the highlighted needs, including:

- Judicial Cases Cross-Check (Evidence standard representation is suitable)

A Judicial Cases Cross-Check system should provide a tool being able to search for case-related information and identify links among cases that are being investigated in other Member States or by JHA agencies and EU bodies.

To facilitate the development of the above solution, a standard representation of the metadata and data of the Evidence should be adopted. In particular the ontology UCO/CASE, dedicated to the digital forensic domain, seems the most promising one to this aim. UCO/CASE, that stands for Unified Cyber Ontology / Cyber-investigation Analysis Standard Expression, provides a structured specification for representing information that are analysed and exchanged during investigations involving digital evidence. To perform digital investigations effectively, there is a pressing need to harmonise how information relevant to cyber-investigations is represented and exchanged. CASE enables the merge of information from different data sources and forensic tool outputs to allow more comprehensive and cohesive analysis. All these metadata represented in a standard format, could be provided to any potential stakeholder using a decentralised repository of metadata along with a suitable level of confidentiality and integrity.

The INSPECTr project (inspectr-project.eu) opted for the open-source UCO/CASE ontology to serve as a standard for interchange, interoperability, and analysis of investigative information.

2)   LEA Capacity Building as a Driver for the Adoption of European Research

Presenter/facilitator: Michael Whelan, experienced cyber analyst at UCD Centre for Cybersecurity and Cybercrime Investigation

The INSPECTr project aims to produce a proof of concept that will demonstrate solutions to many of the issues faced by institutional procedures within law enforcement agencies (LEAs) for combating cybercrime. Unlike other H2020 projects, the results of INSPECTr will be freely available to stakeholders at the end of the project, despite having a low technology readiness level. It is imperative that LEAs fully understand the legal, security and ethical requirements for using disruptive and advanced technologies, particularly with a platform that will provide AI assisted decision making, facilitate intelligence gathering from online data sources and redefine how evidential data is discovered in other jurisdictions and exchanged. However, INSPECTr will also require the support of stakeholders beyond the scope of the project, in order to drive further development and investment towards market-readiness. The development of a robust capacity building program has been included in the project to ensure that LEAs can confidently use the system and that they fully understand both the pitfalls and the potential of the platform.

During our training needs analyses, various European instruments, standards, and priorities are considered, such as CEPOL's EU Strategic Training Needs Assessment, the course development standards established by ECTEG and Europol's Training Competency Framework. With this research and through consultation with internal and external stakeholders, we define the pathways of training for the INSPECTr platform in which we aim to address the various roles in European LEAs and their requirements for the effective delivery and assessment of the course. In keeping with the project's ethics-by-design approach, the training program produced by INSPECTr will have a strong emphasis on security and the fundamental rights of citizens while addressing the gaps in capabilities and training within the EU LEA community. In this paper we describe the process we apply to curriculum design, based on the findings of our research and our continued engagement with LEA and technical partners throughout the life cycle of the project.

3)   INSPECTr: Intelligence Network and Secure Platform for Evidence Correlation and Transfer

Presenter/facilitator: Michael Whelan, experienced cyber analyst at UCD Centre for Cybersecurity and Cybercrime Investigation

Cybercrime is a borderless crime that leverages technology and the internet to exploit businesses, communities and individuals. Law enforcement officers responsible for investigating cybercrime need to be equally able to access cutting edge technology to combat these crimes and to bring down criminal networks. The INSPECTr Project is a solution to many of the issues faced by law enforcement agencies. INSPECTr produces and integrates a range of high-tech approaches, including big data analytics, cognitive machine learning and blockchain technologies into a shared intelligence platform that will improve digital investigations and forensic capabilities, and reduce the complexity and cost of cross-border collaboration. The platform is being designed through extensive collaboration with the law enforcement community, incorporates privacy and ethics by design principles, and takes into account relevant national and international legislation.

After the project, the platform will be freely available to the law enforcement community and adoption will be enhanced through training courses, webinars, and workshops. Exploitation of the project deliverables will also be available to LEAs who wish to further improve the platform, beyond the scope of the project, through additional research and development activities.

Further Opportunities for INSPECTr Dissemination

DFRWSUSA 2022- a premiere research conference for digital forensics in USA, EU and APAC. The CASE ontology, and how that continues to be applied and developed in the INSPECTr project, was the topic presented at the Conference by CCI Senior Analyst, Robbie Dowdall.

EU Cyber Direct Initiative - engages with citizens, research communities and the private sector and civil society organisations through cyber consultations, capacity building activities, policy-oriented workshops and seminars. An overview of the INSPECTr project and the FREETOOL Project were presented by CCI Senior Analyst, Michael O'Callaghan to Policymakers and representatives from the EU and Albania, Bosnia and Herzegovina, Montenegro, North Macedonia, and Serbia.

Project Activities and Events between May 2022 – July 2022

- ➢ INSPECTr Monthly Project Meetings
- ➢ INSPECTr Weekly Technical Meetings
- ➢ INSPECTr LSG Monthly Meetings
- ➢ Ethics Work Package Monthly Meetings
- ➢ EARG Ethics Advisory and Review Group
- ➢ INSPECTr Project General Assembly 2
- ➢ Living Labs Experimentation Phase 3

Regular meeting activities have continued throughout the second quarter of 2022. Included in the regular meetings schedule were two further meetings, key to the project's development.

Living Labs Experimentation - Phase 3

This was conducted in UCD over a three-day period in June 2022. Testing of the current technology using mocked use cases developed by our LEA partners was held on site with support present from the technical development team members to monitor the performance of the evolving technology, troubleshoot and bug fix in a live context. Many issues were addressed during the testing period. Some were tabled to receive attention following the Living Lab as they would require deeper exploration and testing to provide the best solution going forward. Weekly technical meetings continue to develop the technology and prepare for the next Living Lab.

INSPECTr Project General Assembly - 2

The Consortium partners gathered on 20th June 2022 in UCD to review project progress to date and provide status updates within the various project work packages. Management strategies were also discussed to ensure that all technology, reporting, and financial requirements will be on track to be finalised and ready to be delivered by the end of the project. The meeting was extremely productive, being the first physical meeting of the project Consortium since March 2020.

INSPECTr Project Consortium Members representing AGS, BFP, UCD, CNR, EBOS, EPBG, GN, IGPR, ILS, LSP, PSNI, SIREN and Trilateral.

Conferences, Workshops, and Future Events

INSPECTr Consortium Attendance at Conferences and Workshops

UNIL Conference – University of Lausanne	12-05-2022
Artificial Intelligence Hackathon Higher Institute of Electronics of Paris ISEP	13-05-2022 – 15-05-2022
EAFS2022 – Stockholm	30-05-2022 – 03-06-2022
CEPOL Research and Science Conference 2022 MRU, Vilnius: Preparing Law Enforcement for the Digital Age - Vilnius	08-06-2022 – 10-06-2022
DFRWS22 – Attended On-line	11-07-2022 – 14-07-2022
EU Cyber Direct Project – Sarajevo	27-06-2022 – 29-06-2022

| i_LEAD Industry and Research Days - HSD Campus in The Hague | 06-09-2022 – 08-09-2022 |

Closing

We look forward to updating you further in November 2022 with our seventh edition of the INSPECTr Newsletter. In the interim, communications from our readers are welcome and if you wish to contact us or subscribe to our Newsletter you can e-mail us directly at inspectr@ucd.ie. Further information and updates can also be found on our project website https://inspectr-project.eu/.