Intelligence Network & Secure Platform for Evidence Correlation and Transfer

D1.5: INSPECTr evaluation and policy recommendations

Grant Agreement No	833276	**Acronym**	INSPECTr
Full Title	Intelligence Network & Secure Platform for Evidence Correlation and Transfer		
Start Date	01/09/2019	**Duration**	42 months
Project URL	inspectr-project.eu		
Deliverable	D1.5 INSPECTr evaluation and policy recommendations		
Work Package	WP1		
Contractual due date	28/02/2023	**Actual submission date**	08/05/2023
Nature	R	**Dissemination Level**	PU
Lead Beneficiary	PHS		
Responsible Author	Yves Vandermeer		
Contributions from	Ray Genoe (CCI)		

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

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	01/05/23	75%	Initial Deliverable Structure essential content	Ray Genoe
V0.2	04/04/23	98%	Finalisation of the content, conclusion	Yves Vandermeer
V0.3	05/05/23	100%	Minor Refinements	Ray Genoe

Disclaimer

Copyright message

Table of Contents

Table of Figures

Glossary of terms and abbreviations used

Abbreviation/Term	Description
AI	Artificial Intelligence
CASE	Cyber-investigation Analysis Standard Expression
INSPECTr	Intelligence Network & Secure Platform for Evidence Correlation and Transfer
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
LL	Living Lab - A collaborative approach to iterative design and testing with end-users
NLP	Natural Language Processing
PHS	Norwegian Police University College, Politihøgskolen (project partners)
UCD / CCI	UCD Centre for Cybersecurity and Cybercrime Investigation (project coordinators)

1 Introduction

This deliverable details how the project addressed the requirements during the development of the projects and how the technical design combined with the processing of data allows delivering of a product compliant with the EU standards and fulfilling the constraints identified from the point of view of the LEAs for their operational context.

1.1 Mapping INSPECTr Outputs

The purpose of this section is to map INSPECTr Grant Agreement commitments, both within the formal Deliverable and Task description, against the project's respective outputs and work performed.

Table 1: Adherence to INSPECTr GA Deliverable & Tasks Descriptions

INSPECTr GA Component Title	INSPECTr GA Component Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
	Implementation report, including compliance findings and guidelines with respect to the EU forensic policies and legislation, including Ethics, Privacy Protection and Governance. INSPECTr overall assessment and lessons learned	all	The measures implemented to address the requirements and their consequences
TASKS			
T1.4	Evaluation of INSPECTr platform and recommendations for future policy and research	2-4	Chapters 2-3 reflect on the requirements for the INSPECTr platform, how compliant the system is and the impact on end users of the technological results. Chapter 4 lists the policy recommendations produced by the project.
ST1.4.1	Platform Validation and Compliance: perform validation that the platform complies with the defined requirements; analyse the measurements from LL operation	2	Chapter 2 describes the key technical, ethical, legal and social requirements for the platform and how the project has complied with these goals/constraints.

	using quality metrics specified in ST1.1.4		
ST1.4.2	Impact assessment	3	The impact for each of the measures is described in section 3.
ST1.4.3	Lessons learned & future research recommendations; Perform critical and substantiated assessment of what worked well, identifying any issues and problems during implementation	3	Based on the lessons learned during the project, recommendations allowing to improve future work

1.2 Report Structure

The structure of this report is as follows.

Section 2: This section describes the key requirements defined by the project consortium and outlines how the final platform complies with these. The focus is on the technical, ethical, legal and social requirements defined by the LEA stakeholders, the technical team and the ethics manager.

Section 3: This section reflects on the impact of the project, by summarising feedback received from LEA partners and external stakeholders. This impact assessment also provides considerations for future exploitation of the research conducted in the project and how further investment would be welcomed by end-users.

Section 4: Two policy recommendations are presented to the public here. Both recommendations are designed to support future research, through the standardisation of digital evidence and the provision of stronger guidance for LEAs who engage in research projects such as INSPECTr.

2 Platform Validation and Compliance

The Technical Platform developed by the project, with considerations of the specificity of the end-users and the sensibility of the handled data to be validated and checked for its compliance with technical, ethical, legal, and social requirements. Moreover, the final version must then be projected in an LEA operational environment. This was achieved by involving the LEA steering group, the LEA end-users as partners in the project and the operational experience from PHS in the technical developments.

2.1 Technical Requirements

When validating the technical requirements, the platform was considered as a set of different components and, as a whole.

1. The basic functionalities must realise INSPECTr objectives. The project results include:
    - Integrated digital forensic and investigative tools
    - Investigative assistance from AI/ML models for managing big data sets
    - Robust and scalable storage of digital evidence
    - Ability to access and analyse existing LEA data external to the system
    - A facility for evidence discovery between nodes
2. The platform should be a desirable solution for LEA and acceptable from a societal perspective with respect to fundamental rights. INSPECTr addresses this as follows:
    - LEAs were included in the design and testing phases, so anticipating potential constraints raised by the operational environment, actual guidelines, and standards.
    - This was in addition to strong technical, legal, social, and ethical oversight throughout
3. The final platform must be freely available to exploit by EU-LEAs, and hardware requirements must be cost-effective. To ensure this:
    - The project chose a minimal hardware configuration to test the technologies developed on low-cost solutions for LEAs, while ensuring scalability on demand
    - All the technology developed or included in the project is freely available to LEA for further exploitation. This is facilitated by compliance with the FOSS (Free and Open Source Software)
4. The platform should be highly modular, allowing future technological advances to be incorporated. This has been assured by:
    - The containerisation of most INSPECTr tools
    - The standardisation of evidence through the adoption of a leading evidence ontology, will promote further development and interoperability
5. The system should be easily deployable for providing rapid installation/updates to end-users. INSPECTr achieves this since:
    - Updates can be quickly deployed when issues are identified and addressed
    - New tools or improved technologies can be rapidly deployed and integrated when available in the future

6. The interface should be highly intuitive, requiring minimal training to use. INSPECTr has achieved this in several ways:
    - The platform provides a unique single interface to various investigative tools, with similar report dashboards and menu systems.
    - The case management system has been designed and refined through consultation with end-users
    - Monitored experimentation sessions have helped inform the design of training material. In addition to ethical information, the teaching material focuses on the difficulties encountered by testers, i.e., the less-intuitive aspects of the platform

7. Operational sensitivity should be respected by the system. INSPECTr provides various solutions to this concern.
    - Access to data is restricted within a node since it provides a multi-tenancy of users. This means that a user must be explicitly granted access to existing case data.
    - Access to data is restricted between nodes (e.g., jurisdictions), currently and without exception. In future, this can be made possible in cases where a joint investigation is approved, or the current legal instruments (MLAT/EIO) permit the exchange of data.
    - "Data discovery" between nodes can be used to indicate that one node may contain similar evidential items to another without seeing investigative data from the other node. This is strictly controlled through legal and administrative controls and further ensured by using the traffic light protocol throughout the system. The traffic light protocol is used for classifying all evidential data and subsequent sharing or processing rights.

2.2 Ethical, Legal and Social Requirements

There was a dedicated work package for ethical and privacy issues in the INSPECTr project, and the project teams included expertise on both social and legal aspects of privacy as well as engineers with privacy and information security expertise.

1. Access **to Tools:** Certain policies or regulations within an LEA jurisdiction, or even with an LEA unit, may not permit the use of some tools that are used in other jurisdictions/units. Controls should be in place to limit the use of all tools.
    - All tools in INSPECTr are disabled by default
    - Only an administrator - for example, a senior officer - may enable these for investigators
    - Detailed information about the tools provided in the interface and in the accompanying training programme for administrators and investigators

2. Limitations **of Tools:** Certain tools may not be as effective as the user perceives them to be. For example, the legal reasoning engine is subject to the varied nature of legal systems and the evolving nature of law. Users should be made clearly aware of this. As a solution:
    - The limitations of certain tools are provided in the interface and the accompanying training materials.
    - The classification models provided to LEAs include information on biases with adjustment measures taken depending on the biases. Newer models can be provided, due to the platform's modular nature. LEAs can also re-train some models, if biases are found or regional aspects (e.g., dialect) necessitate fine-tuning them.

- Access to certain tools can be restricted if they are not performing correctly or are outdated. The method described in the "Access to Tools" point above enables an administrator to easily disable any tool in INSPECTr.
- A "human in the loop" solution is provided for critical decisions concerning the results of AI classification or the exchange of information between nodes

3. Data Minimisation should be respected throughout. To address this:
   - Data is only retained when necessary for the purpose of investigative integrity
   - Online investigation tools can only request/retain data that is pertinent to an investigation
   - Data can be deleted manually or automatically, with prompts to the investigator to proactively retain the data, if appropriate and required under local circumstances
   - The training curriculum produced includes a dedicated module on ethical, legal, and societal issues, where privacy is a key focus.

4. **System** misuse should be difficult throughout. INSPECTr addresses this with the following measures:
   - User and data management procedures ensure controlled data access
   - Chain of custody and chain of evidence records prevent manipulation of evidence
   - Evidence discovery can only happen under strictly controlled legal rules and sharing agreements, all of which must be validated by an administrator
   - Evidence discovery and AI tools can only be used on investigative data within INSPECTr, which has provenance records,
   - The use of any tool in INSPECTr can be traced back to an investigator since every action is logged by default, thereby providing retrospective accountability

5. **Technical and organisational measures** were implemented to safeguard data subjects/research participants and **security measures** to prevent unauthorised access to personal data. INSPECTr addresses this with the following measures:
   - The ethical monitoring of the development and technical implementation, including a deliverable on how the incidental findings are handled and risks assessed.
   - The development of a full Case Management System using an innovative security mechanism based on tokens and users' wallets that can secure features, investigation and data based on rights granted by a supervisor.
   - The embedding of a blockchain mechanism tracing all actions, including the grant and modification of users' rights.
   - The limitation, by design of the components to avoid the risk of mass surveillance.

6. Precautions to **eliminate or** mitigate **potential AI or Data Mining algorithmic** biases. This was addressed by the selection of compliant AI tools and validated during the ethical monitoring of the development, as during the last living lab.

3 Impact Assessment and Future Research

The findings presented here are the product of continuous consultation with LEAs throughout the project. The findings are related to the feedback received about the technologies presented, comparisons to existing technology, their suitability for local deployments, future exploitation, and ethical, legal and security aspects of the technology as a whole. This section summarises these findings.

LEA engagement has played a vital role in developing the INSPECTr platform. The consortium partners formed an LEA steering group and helped to design and test the technologies developed by the technical team from the first project meeting and throughout the project to the last technical test. The LEA partners also helped design and develop the scenarios to test the platform to ensure that the technology would be grounded in real-world scenarios they encounter in their daily work.

This was reflected in the feedback received from external stakeholders, and the majority of concerns expressed during platform demonstrations seem to have been addressed by the project. The project's dissemination team ensured that, as well as raising awareness about the technology being developed, external stakeholders were able to provide feedback throughout the project. This would ensure that a wide range of LEAs and other stakeholders would be able to raise issues from all corners of the European Union to ultimately provide a platform that would meet the needs of all LEAs.

3.1 Technology Review

While the platform has a low technology-readiness level (TRL 6) and has been built as a proof-of-concept solution for addressing many of the challenges faced by law enforcement, many of the techniques demonstrated have been welcomed by LEAs and the project consortium has been encouraged to pursue further development of the platform.

The integrated tools and parsers have received the most attention due to their low cost, simplicity of use, and quality of outputs produced. In addition to the ability to parse commercial tool reports, the suite of AI tools has also been well received. Using such a wide range of tools in a single environment to digital forensic and intelligence gathering tools, is something that LEAs are not accustomed to, but identified as a positive operational asset in future, especially due to the minimal training required.

When processing large binary files and returning large forensic reports, the platform currently suffers large data transfer overheads when passing result sets from one subsystem to another. Optimisation of data transfer between INSPECTr node subsystems is an area of future work that will greatly impact the processing speed within the platform.

3.2 Administrative Controls

One of the concerns raised by LEAs centres on the diversity of tools available to investigators. While some welcomed the array of digital forensic (DF), open-source intelligence (OSINT) gathering tools, website preservation tools (web-scraper) and Artificial Intelligence and Machine Learning (AI/ML) tools, others expressed concerns about local policies and legislative restrictions in their jurisdiction.

For example, many LEAs expressed concern over the use of AI/ML tools, particularly with uncertainty surrounding the content of the proposed EU Artificial Intelligence Act, which is being prepared and should be in force later this year (2023).

INSPECTr addresses all of these concerns through the implementation of administrative controls on the entire suite of tools that can be made available to investigators via the platform's case management system (CMS). The toolbox is made available via the docker orchestrator, which allows users to create groups of investigators and assign tools to them that comply with local policies. This means that the system has the flexibility to both impose restrictions and keep up with legislative or policy changes in the future. The system is also multi-tenancy; multiple different "organisations" (or groups of users) can coexist with different capabilities.

The administrative controls are intended to be used by senior management and are transparent to investigators using the CMS. In other words, only the tools that have been made available are apparent in the CMS. As an added benefit, tools that have proven to be unstable or unsuitable for use can be disabled by the administrator, while new versions of tools can also be enabled. All LEAs considered this to be a very valuable asset to the implementation of the INSPECTr platform.

3.3 Ethical and Security Implications

The project has involved strong ethical oversight from the outset and discussions with LEA partners and technology providers have helped create the platform under strict ethics-by-design and privacy-by-design principles. This is not only to ensure that misuse of the platform can be avoided but also to safeguard LEA from criticism or legislative measures taken against them for using the platform's technologies.

Special attention was given to the AI/ML tools developed during the project, since there is a potential for misuse or unsuitable outcomes from such tools. For example, facial identification, gender classification and sentiment analysis are contrary to modern concepts of responsible AI, due to the importance of data privacy, the fluidity of gender, and the lack of accuracy in sentiment analysis with AI systems. In addition, AI systems should be assistive technologies, and not decision makers in their own right, that present suggestions to an investigator rather than eliminating data from an investigative line of enquiry. Furthermore, the importance of avoiding bias that the trained models have incorporated means that greater scrutiny is required when using pre-trained models and an awareness of the quality and provenance of the datasets used to train them is also of great importance.

LEAs welcomed several aspects that were employed by the INSPECTr project with regards to ethical AI. In addition to the administrative controls provided, which will allow inappropriate/unsuitable tools to be blocked from use, the AI toolbox provides the ability to retrain models on regional needs; for example, fine-tuning the Natural Language Processing models to accommodate local dialects. To further our accepted approach to ethical AI, all of the image classification tools have been designed to prioritise data, rather than exclude or make decisions on what is important or not. This means that an investigator will always have to view the data, but the prioritisation of the data will make their job less arduous.

Techniques such as the prioritisation of data can ease the time and psychological impact that visually processing large volumes of material can have on investigators. Child Sexual Abuse (CSA) investigations in particular will greatly benefit from the use of AI tools and digital forensic units often find that these are within the majority of investigations encountered. These types of investigations also cause large backlogs, due to the volume of media and devices that are often seized. INSPECTr provides triage tools, which are designed to address this backlog through fast classification of devices,

for prioritisation for full forensic analysis. However, the added benefit is that non-technical LEA officers, and those without sensitivity training to handle CSA material, can also use the tool without being exposed to the findings. This is just another technique to the benefit of LEA that is seen as one of the major successes of the platform.

One of the most interesting and potentially disruptive technologies developed by the INSPECTr project is the publish/subscribe approach to evidence discovery between LEAs, an area typically fraught with concerns around operational security, misuse and data privacy. LEAs welcomed the concept that legal agreements and administrative oversight would control the way questions can be asked (and answered) between LEA in separate jurisdictions about the evidence in their cases. They also welcomed the fact that this can only be done from within the context of an ongoing investigation; i.e., an officer cannot, for example, check if their phone number is under investigation in another jurisdiction as the data must originate from an ongoing investigation and various logging activities can trace the origin of the data.

3.4 Accessing Technologies and the Network

While the publish/subscribe system is of great interest to LEAs, the manner in which it is facilitated requires much more work. LEAs wondered how many would be able to connect and who would have access to the central node and, therefore, access to all of the requests. In truth, the system itself cannot handle that many connections in its current form and the e-codex team are working on a better approach to ensuring that a) more nodes can be connected and that b) larger data transfers can be facilitated across the network. Upon further development, it will be possible to realise a pan-European network of nodes, with a central node hosted by an authority who can initiate joint investigations and full data exchange.

Another issue raised by consortium LEAs was about the complexity of installing an INSPECTr node. Quite often the LEA officers have limited IT support and conduct many of the network and system administration tasks themselves. During the training course for node installation, and the feedback received afterwards, it became clear that the technical team would have to help some partners with their installations and explain the manner in which networks should be secured to protect local infrastructure and the node itself. When considering the sustainability of a mature version of the platform, it may be prudent to consider a service-level agreement that could include installation and maintenance of the nodes and could potentially be costed to also include future developments. This was the suggestion from a few LEAs during recent discussions around the future of the platform.

On a more positive note, almost all LEAs were impressed with the deployment strategy of the nodes, once the more complex issue of setting up the infrastructure was overcome. The way that tools can be updated to address bug fixes or feature requests was highly regarded by most LEA stakeholders. Essentially a tool developer can make a change to a tool and the changes can be reflected on all nodes in the network within minutes. This is not just for the investigative tools but also applies to almost all of the services on an INSPECTr node. This manner of agile development is rarely seen with commercial tool vendors. Many LEAs liken it to the interactive approach seen on the FREETOOL project, where end-users can directly report issues to tool developers.

3.5 Benefits compared to current technologies

In addition to the method of deployment and the future possibilities of having direct access to an agile development team, LEAs were impressed with the considerations that have gone into the platform as a whole. We firmly believe that this was made possible through constant consultation within the consortium, LEA steering of the design and strong ethical governance over the developed technologies.

We have also strived to ensure Free and Open-Source Software (FOSS) compliance from the outset of the project, rejecting licensed technology that was brought to the table by some industry partners in favour of using open-source tools where possible and developing our own when not. This has resulted in a license-free solution for LEAs that is very attractive to them.

No LEA that was consulted during the project was able to mention a current platform that boasted the wide-ranging technology provided by INSPECTr and the publish-subscribe system is widely regarded as the first of its kind. While there was some apprehension about the potential for adopting this approach, almost all LEAs could see the benefit if the platform could be matured towards operational use, and with the backing of international police organisations.

3.6 Future development of the platform

As a proof-of-concept TRL-6 platform, more work is required before INSPECTr can be fully operational for LEAs to use in their investigative workflows. Feedback from LEA was largely positive when asked if the project should merit future investment. Many said it was an operational asset that many EU agencies would benefit from and would like to see the platform reach full maturity. In addition to the current LEA partners, a number of other LEAs have shown an interest in being part of future research into the platform and its technologies.

From the outset, the consortium has strived to promote the work conducted in the project to the relevant stakeholders. However, we have also promoted open research practices so that other researchers who share the same vision can join the development team in future. Furthermore, during discussions held at our dissemination activities, we considered various approaches to further development. LEA was keen to see their budgetary restrictions considered in the final platform offerings, so further funding from the European Commission would likely be required to ensure that the results could be cost-effective.

Many LEAs highlighted the lack of similar platforms for conducting a wide range of investigative processes while combining AI/ML and analytic tools. Furthermore, discovering evidence with other jurisdictions is seen as a game-changer in the field. Therefore, some LEAs would like to see the platform developed as a whole and sustainability considerations to be included in future. It remains to be seen if a funding programme such as Horizon Europe would suit the development of the entire platform to TRL 8 or TRL 9, to ensure suitability for operational use. This would take the support of a large number of LEAs, which the consortium believes can currently be counted on.

4 Policy Recommendations

The INSPECTr project has consulted LEA from the outset, and they have been involved in every aspect of the iterative design and testing phases within our "Living Lab" ecosystem. As the project developed, the consortium was able to identify technical solutions to some of their challenges.

However, some issues faced could be addressed with a stronger approach from the EU Commission. As a result, we present two policy recommendations which we hope will shape the criteria for future research in the area and a more harmonised approach to LEA investigative technologies.

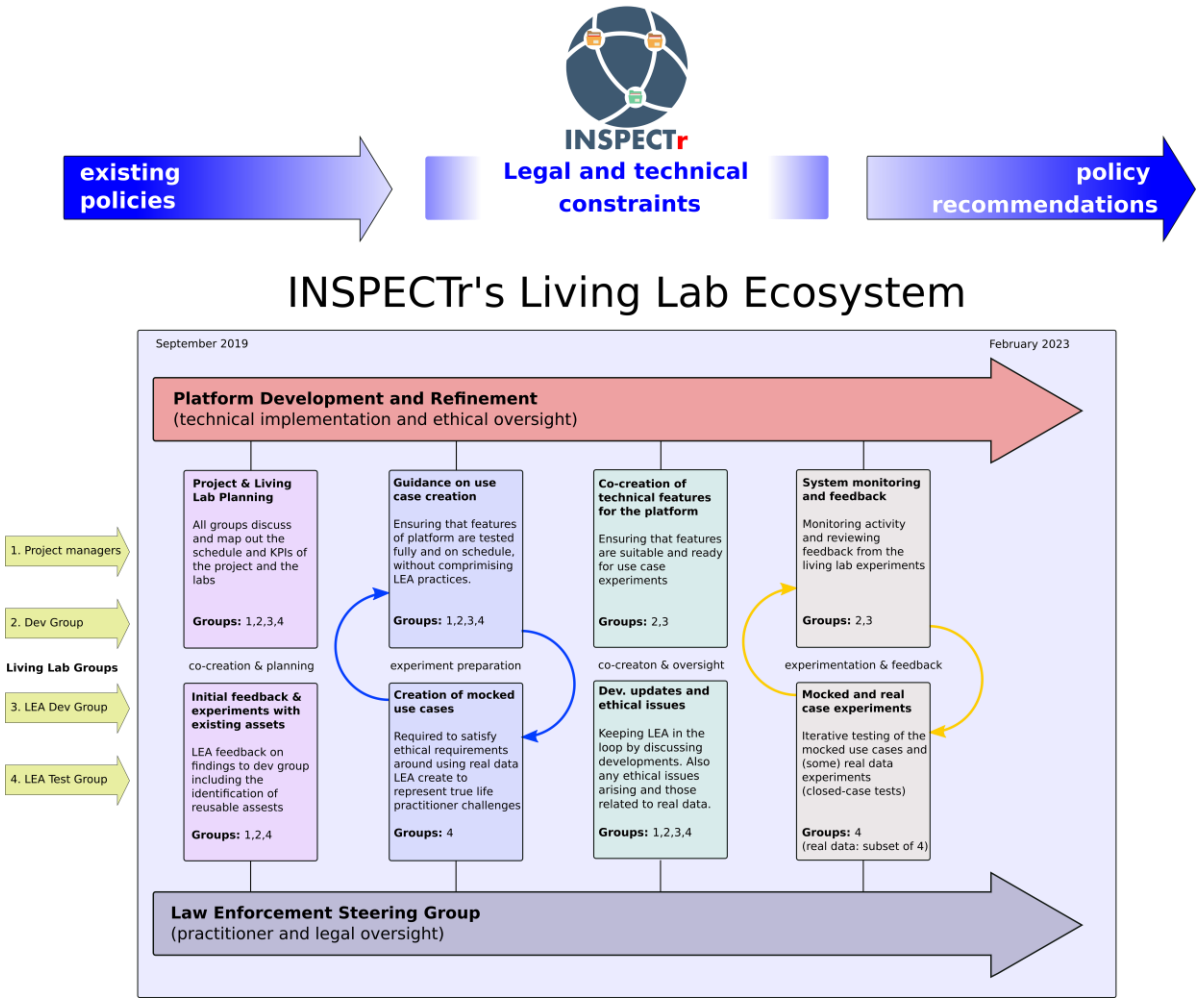

Figure 1: INSPECTr's Living Lab Ecosystem

4.1 Digital Forensic Investigations and the Standardisation of Digital Evidence

Our recommendations include:

- The adoption of the Cyber-Investigation Analysis Standard Expression (CASE) ontology as the standard output format for all investigative tools
- Encouraging the use of *CASE* output in all vendors' developed investigative tools.

- The use of CASE output for all investigative tools and platforms developed for law enforcement using EU funding.

The overarching objective of INSPECTr is to support more effective prediction, detection, and management of cybercrime by developing an online solution for law enforcement, within which investigative data can be securely shared and analysed. Each law enforcement agency will have access to a national-level INSPECTr node, which will feed into a Pan-European platform where digital evidence from multiple agencies and/or jurisdictions can be cross-correlated. This will result in faster and more accurate investigations and aid in identifying previously uncovered criminal actors and actions. The INSPECTr platform includes a range of cutting-edge analytic tools facilitating both reactive and preventative policing and will be freely available to law enforcement. Ultimately, the solution will support collaboration and partnership between law enforcement agencies, speed up digital investigation processes, and enhance law enforcement's capability and capacity to use sophisticated technologies to aid their work competently. The platform incorporates privacy and ethics by design and has considered both national and international legislation.

As the key aim of INSPECTr is to simplify digital investigations, it was essential that features, tools, and processes were complementary with each other. Moreover, the free nature of the platform means that whatever solutions are deployed, they must be easily maintained. The challenge with this is the ever-increasing range of digital evidence, from laptops, mobile devices, CCTV cameras etc., each having a different output format. In addition, law enforcement agencies will have their preferred commercial and/or free digital tools for conducting forensic analysis. The wide range of evidence, coupled with the wide range of tools used to extract that evidence, leads to a multiplicity of formats in output. Given these disparate outputs, cross-exhibit analysis becomes increasingly difficult. Therefore, a common standard was necessary for interoperability. As the Cyber-investigation Analysis Standard Expression (CASE), had already been integrated into several previous projects funded by the European Commission[1], the decision was made to include and further enhance CASE as part of the INSPECTr project.

The Cyber-investigation Analysis Standard Expression (CASE)

The growing number of investigations involving digital evidence from various data sources demands a standard way to represent, analyse, and exchange pertinent information. The Cyber-investigation Analysis Standard Expression (CASE)[2] is a community-developed evolving standard that provides a structured (ontology-based) specification for representing information commonly analysed and exchanged by people and systems during digital evidence investigations. It is an international standard supporting automated combination, validation, and analysis of cyber investigation information.

From a technical perspective, interoperability of intra-jurisdictional systems calls for the development of standards in how data are represented and exchanged, and there are obvious benefits to having a standard output format:

- **Fostering interoperability.** Enables the exchange of cyber-investigation information between tools, organisations, and countries. For instance, organisations involved in joint investigations can easily share information using CASE.

[1] https://www.e-codex.eu/

[2] https://caseontology.org/ontology/intro.html

- **Tool Validation.** Expectations of a common output from different tools can support validation, as results can be compared and evaluated for their completeness and correctness/accuracy.

The primary motivation for CASE is to lessen the analytic burden of cyber investigators by providing a common language to support automated interoperability across a variety of information sources to facilitate the analysis and exploration of investigative questions. However, in addition to advancing the efficient and accurate exchange of cyber-investigation information between tools and organisations, CASE ensures that analysis results can be traced back to their source(s), keeping track of when, where and who used which tools to perform investigative actions on data sources.

Therefore, **CASE** also brings additional value to the investigative process by:

- **Strengthening Chain of Evidence.** A fundamental requirement in digital forensics is the consistency and reliability of evidence provenance as it is exchanged and processed. CASE provides better authenticity and trustworthiness with regard to Chain of Evidence and Chain of Custody relying on a clear representation of the Chain of Evidence (provenance) and the Chain of Custody.

- **Advanced Analysis.** CASE enables more advanced and comprehensive correlation and analysis. In addition to searching for specific keywords or characteristics within a single case or across multiple cases, having a structured representation of cyber-investigation information allows more sophisticated processing such as data mining, or NLP techniques. This can help, for instance, to overcome the linkage blindness issue that is the failure to recognise a pattern that links one crime to another, such as crimes committed by the same offender in different jurisdictions.

The INSPECTr project has significantly contributed to the development and implementation of the CASE standard in operational settings. However, it was necessary to parse existing tools integrated into INSPECTr to conform with the CASE data structure. This was to circumvent the issue of multiple output formats, thereby allowing investigators to continue using their tool of choice. **The introduction of CASE as a standard output format would negate the need for additional engineering of developed tools, as they would already be fit for purpose.**

As CASE is a community-based initiative, all improvements to the standard are fed back to the CASE community, so that the entire membership benefits as the standard is continually evolving.

Policy-Relevant Findings

The EU has several instruments under which law enforcement, cybercrime/cybersecurity and Information Technology are addressed:

- The EU Security Union Strategy[3] outlines key challenges threatening the security of the Union, laying out strategic priorities for action. These priorities include enhancing law enforcement capacity in digital investigations, the creation of a Joint Cyber Unit to address the growing cyber threat, and a review of the EU rules against cybercrime. Further support to law enforcement will be provided by the EU's Joint Research Centre, which will provide stringent scientific evaluations and testing methods for any research activity targeted towards law enforcement.

- The European Cybercrime Centre (EC3)[4] at Europol was established in 2013 to strengthen the law enforcement response to cybercrime and provides highly specialised technical and digital forensic support capabilities to investigations and operations. More latterly, the recently formed Europol Innovation Lab[5] has an EU mandate to support the law enforcement community in the area of technology innovation.

- The European Union Agency for Cybersecurity (ENISA)[6] is the entity responsible for supporting the development of cybersecurity capabilities across the Member States. A current ENISA initiative is the creation of an EU cybersecurity certification framework to advance the security of software developed in the EU.

- The European Anti-Cybercrime Technology Development Association (EACTDA)[7] is a brand-new EU funding initiative designed to support the development of technological solutions for European Law Enforcement Agencies and Forensic Laboratories.

- Finally, the EU Rolling Plan for ICT Standardisation[8] specifically looks at how standardisation and interoperability can support EU policy goals.

Committing to the deployment of the CASE standard will support the work of the above instruments in a variety of ways:

- Enhancing cooperation between Member States' law enforcement agencies through the harmonisation of information formats.
- Improved cross-border acceptance and exchange of court-proof evidence.
- A common tool validation mechanism.
- The ability to validate/evaluate LEA commercial and non-commercial tools.
- The ability to harmonise research outputs and commercial tool outputs into common platforms.
- The ability to describe public datasets for EU research (to validate research outputs.
- Supporting EU standardisation policy.
- Offering an additional mechanism/requirement for secure software certification.

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0605&from=EN

[4] https://www.europol.europa.eu/about-europol/european-cybercrime-centre-ec3

[5] https://www.europol.europa.eu/operations-services-and-innovation/innovation-lab

[6] https://www.enisa.europa.eu/

[7] https://www.eactda.eu/

[8] https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation

Challenges and Opportunities

CASE Community - The maintenance and further development of the CASE standard is managed and run by a community of volunteers and is an open-source initiative. There is obviously a potential risk in dependency upon a community that relies mainly on donations to support its efforts. Furthermore, the CASE standard is still relatively immature. However, the open-source nature of the initiative means that development work undertaken is undertaken by members. Therefore, as the community grows and the CASE standard is more widely adopted, there will be an exponential increase in development work, leading to greater enhancements and improvements.

The INSPECTr project has significantly contributed to the improvement of the CASE standard. The project has developed parsers for commercial tools (UFED, AXIOM, XAMN and OXYGEN) in order to convert their output (XML) into CASE. During these activities, the developers engaged actively with the CASE community, highlighting issues, and proposing solutions and improvements. Developments under INSPECTr have been included in the more recent version of the CASE ontology and have greatly enhanced its maturity. It is also worth mentioning that since December 2021, the CASE project is part of the Linux Foundation, and its membership includes the Netherlands Forensic Institute, the National Institute for Standards and Technology (NIST) and Europol EC3. Securing the initiative through targeted funding would certainly assuage concerns around the long-term viability of the CASE community.

Vendor Resistance – One of the objectives of the EU Security Union Strategy is to create closer cooperation between players in the public and private sector. Requesting vendors to significantly revise their development procedures is a tough ask. There would need to be a concerted effort to 'encourage' vendors to commit to such a plan. However, EU strategies, such as the ENISA Cybersecurity Certification scheme might be used as leverage to encourage vendor compliance.

Despite the challenges, standardisation is key to:

- 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

4.2 Data Protection Regulations

Data protection policy recommendation:

- a cross-European organisation, such as the European Data Protection Board, should clarify the legal framework applicable to research using personal data, especially for use of highly sensitive law enforcement data in security research. There is substantial fragmentation in the implementation of Member State law creating compliance risks and it is imperative to clarify where processing by LEAs for research purposes could fall into either the General Data Protection Regulation (2016/279, 'GDPR') or Law Enforcement Directive (2016/680, 'LED') regime[9].

[9] A more detailed analysis of the issues and policy recommendations presented here is being prepared for submission to an academic journal by Trilateral Research.

European Commission projects in the security domain are encouraged to include law enforcement agencies (LEAs), and to use 'real' data in from closed cases projects (data from ongoing cases likely poses an overwhelming ethics risk) for the testing (and potential training) of new technologies, including artificial intelligence (AI) technologies. However, the data protection framework applicable to the processing of closed case data for research purposes is not at all clear.

The GDPR is intended to regulate the processing of personal data in 'general' situations, and the LED is intended to govern such processing for law enforcement purposes. Normally, scientific research would come under the GDPR, and there is a 'special regime' within the GDPR to regulate this. However, in security research, processing can take place for both research and law enforcement purposes simultaneously. This leads to a question about which regime the processing should be regulated by.

The lack of clarity could lead to organisations (primarily LEAs) processing data under one legal framework, whereas they should be processing under a different framework (and could also lead to 'shopping' for a governance regime that is least onerous). For individual organisations, this lack of clarity can affect the ability of organisations to comply with data protection legislation when they participate in security research projects. For collaborative research projects, it can create difficulties for collaboration between LEAs as research partners where data is exchanged.

Policy-Relevant Findings

This policy brief addresses the EU's data protection acquis and the regulatory approach to scientific research in the law enforcement domain. Whilst the development of the GDPR and LED clarified numerous issues relevant to data protection, these sister pieces of legislation have also created some complexities for scientific research that are yet to be clarified.

The nature of scientific research under data protection is a topic that has received some significant attention, though this is still insufficient. For example, the European Data Protection Supervisor (EDPS) has written on data protection issues relevant to scientific research, but noted several areas that need further consideration by researchers, including as part of Horizon 2020 and Horizon Europe projects[10].

All research projects funded by the Horizon framework programmes are subject to these difficulties. However, research with LEAs has the added complexity of an alternative legal regime that could equally apply to the same activities.

During the INSPECTr project, it has become clear that the situation with respect to the appropriateness of different legal regimes for processing these types of data in research projects is very unclear. Indeed, research in INSPECTr has demonstrated that it could be possible to use both legal regimes, but that the GDPR regime should be favoured. This lack of clarity is potentially problematic, as data controllers could process LEA data for research activities under one legal regime and regulators could take an entirely different view.

Should this policy recommendation from the INSPECTr project be taken up, it would go a significant way to clarifying an often misunderstood area of data protection law. Clarity on this matter would provide a clear foundation not only to other Horizon research projects working with LEAs, but also facilitating collaborations between LEAs and other researchers. LEAs are increasingly collaborating

[10] Wojciech Wiewiorówski, "A Preliminary Opinion on data protection and scientific research", *EDPS*, 6 January 2020, p.24-26.

with researchers to develop specific technologies for their needs that are unavailable on the open market. Although INSPECTr and many similar projects have a technology focus, this issue also affects other areas of research, such as criminology, social sciences, and humanities, where research is of increasing importance in order to understand the needs of LEA stakeholders.

As noted, the EDPS has explored the topic of scientific research and data protection, as have some academic research articles. Whilst some debates are ongoing, there is a general convergence toward the position outlined by the EDPS.

The GDPR regime regulates scientific research through requiring one of the six legal bases for the processing of personal data to apply; which of these is most appropriate depends on the circumstances of the processing. Yet, the EDPS is non-committal on the ability of the 'public task' legal basis to apply to scientific research[11], despite this being used by many public authorities, including universities, publicly funded research institutes, and some LEAs.

In INSPECTr, legal research has identified that there is a variety of different approaches in national law to the inclusion of research as an activity of public authorities, such as LEAs. For example, scientific research is one of the public functions of Romanian LEAs. Yet, this is not the case in Greece, where scientific research is not a public function of LEAs. Such LEAs need to rely on a different legal basis, such as legitimate interest. As such, there is clear fragmentation in the national legal approaches to the regulation of scientific research by LEAs.

However, research with law enforcement agencies has the added complexity of interaction between the GDPR, which regulates 'standard' scientific research, and the LED that governs the activities of LEAs. Where LEAs process personal data for law enforcement purposes, this is regulated under the Member State laws emanating from the LED. According to Art.1(1) of the LED, law enforcement purposes can include:

> *'the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.'*

Arguably, scientific research can contribute to crime 'prevention'. Further, the LED provides for scientific research to take place under this regime through member state law. Art.4(3) states:

> *'Processing by the same or another controller may include archiving in the public interest, scientific, statistical or historical use, for the purposes set out in Article 1(1), subject to appropriate safeguards for the rights and freedoms of data subjects'.*

Research in INSPECTr has noted that not only is there fragmentation in terms of how Member State laws affect the legal bases available to LEAs and other public authorities, but there are some countries that have made use of the provisions in the LED to allow for scientific research to take place under this regime; for example, the Irish Data Protection Act 2018.

The fragmentation identified can lead to issues for LEAs engaging in collaborative research projects:

- Firstly, as different legal bases are available to different LEAs, this can place different requirements onto them. Whilst all persons and organisations processing personal data will need to consider the

[11] Wojciech Wiewiorówski, "A Preliminary Opinion on data protection and scientific research", *EDPS*, 6 January 2020, p.23.

data protection implications of doing so, an LEA who can rely on a 'public task' legal basis would not need to identify and balance their interests in engaging in research against the rights and freedoms of the data-subjects, as is necessary for LEAs who rely on the legitimate interest legal basis. It is possible that one LEA who can rely upon the public task legal basis could proceed with some processing, whilst another LEA might be unable to rely upon legitimate interest if they fail to adequately balance their interests with those of the data-subjects (of course processing that would fail the balancing test should often not take place, even if public task can be relied upon). Thus, this fragmentation can lead to different capacities for LEAs to engage in research, with the potential implication that different LEAs will be unable to collaborate in some sensitive research activities.

- Secondly, although the sharing of LEA data can be difficult and onerous from the data protection perspective, having LEAs collaborate and share data for research purposes can bring great benefits to research projects. Where LEAs have different legal bases for processing personal data from other LEAs or other research partners (e.g., universities), this could potentially mean that data sharing becomes more difficult than it would otherwise be as joint-controller and data sharing agreements are easiest to implement where the different parties can rely upon the same legal basis and are governed by the same legal regime. Controller-Processor Agreements where the legal regime and legal basis is set by the Controller could be implemented in some circumstances but might not be appropriate for all research activities.

- Thirdly, it is possible that an LEA could take a particular view on which legal basis, or which legal regime is most appropriate. However, their supervisory authority, or a court, might take a different view, leading to an LEA being in violation of its data protection obligations, despite acting in good faith and in the belief that it was acting lawfully. Thus, this fragmentation and lack of clarity presents compliance risks for LEAs themselves.

Consequently, research in INSPECTr identifies that it would be most appropriate for parties in collaborative research projects to process personal data under the same legal regime and, ideally, rely upon the same legal basis. This would give clarity to the extent of the data processing that is applicable and enable consistent oversight of data processing.

Challenges and Opportunities

The GDPR contains a 'special regime' for the regulation of scientific research. Art.89 allows for Member States to implement national legislation that would allow researchers to derogate from certain data protection obligations. Further, Art.9(2)(j) allows for an exemption to the general prohibition to the processing of special category personal data to be created in national law for scientific research purposes. Allowing Member States to regulate certain aspects of EU law on the national level is often done to enable different national sensitivities to be taken account of where a single approach on the EU level would be inappropriate. As such, some fragmentation of the GDPR regimes applicable to scientific research is intended by design.

This could mean that conformity across the EU in terms of scientific research, and specifically on law enforcement research might not be desirable from an EU-level perspective. However, clarity on how the 'special regime' is intended to function on the EU-level would still be beneficial, even if it is to ensure that the differences between how each country approaches the issue is better understood.

The opportunity here is that greater clarity on how LEAs can approach data protection for their research activities can not only facilitate smoother research where LEAs are already engaged in research activities, but it could facilitate more research between and within LEAs and with other researchers. There is a societal imperative for LEAs to optimise the way that they fight crime, in light

of the increasing complexities of, and amounts of data collected during, criminal investigations, and the speed at which they are expected to progress. Research into new technologies to assist LEAs contributes to this significantly. Ultimately, better technological capabilities can help to enhance investigations, thereby contributing to public safety and security.

The conclusion of this research conducted in INSPECTr recommends that clarity on this situation should come from a cross-European body. The European Parliament could pass new legislation, but there is limited political enthusiasm for re-opening the data protection acquis. Professional associations could offer guidance to their members, but there is limited scope for this to be accepted across many organisations that might not have association members present; although, groups such as the EUROPOL Data Protection Experts Network could make a significant contribution to creating clarity if they were able to agree on a position. Individual data protection authorities could provide guidance, but this is likely only to be accepted within their jurisdiction. Whilst the EDPS has provided an influential opinion on scientific research, and a final opinion would be welcome, the authority of the EDPS is, in strictly legal terms, limited to EU institutions. As such, the most appropriate body to provide guidance that would apply across the EU, and have significant authority behind it, would be the European Data Protection Board (EDPB) who could incorporate this into their wider guidance on data protection in research.

Having said this, if the EDPB were to accept the recommendation of this paper and provide guidance that the GDPR should be seen as the primary legal regime for regulating the processing of personal data for LEA research, this should be presented in a very specific way that does not exclude alternative approaches. Firstly, the LED has been drafted to allow for research to take place under its regime and it is not the place of the EDPB to prevent this. Secondly, such guidance could be seen as posing a challenge where Member State legislatures have already provided for scientific research to take place under the LED (e.g., Ireland); they might not be inclined to accept the views of the EDPB on revising their national legislation. As such, any guidance should specifically note that whilst there are situations where processing personal data for scientific research purposes might be appropriate under the LED, this should only be in exceptional circumstances.

Policy Implications

The drafting of guidance by the EDPB is appropriate as the fragmentation between states is leading to challenges that are difficult and complex to resolve. The divergences in research approaches between different LEAs across the EU are increasing and becoming entrenched. Current guidance from the EDPS does provide some benefit, but greater clarity is needed so that differences can more easily be overcome by LEAs and their research partners.

Conclusions

LEAs in Horizon research projects are encouraged to use case data during research so that new technologies, including AI technologies, can be tested (and potentially trained). Yet, the required processing of personal data needs to take place under an appropriate legal regime. The legal research conducted in INSPECTr demonstrates that the existing lack of clarity about which legal regime should apply to scientific research, and which legal bases are appropriate, along with fragmentation across Member States poses challenges for LEAs who engage in research. Some LEAs can, and do, engage in research processing personal data under the LED legal regime, whilst some follow the GDPR regime. Such fragmentation could prevent some research activities taking place and can make others much more difficult. This prevents the opportunities of collaborative research being seized, meaning that

the results of research cannot be harnessed for public safety and public security as easily. Clear guidance on how data protection legislation should apply to scientific research by LEAs from a body such as the EDPB would be greatly beneficial before such processes become entrenched.

5 Conclusion

The combination of the efforts on ethical and technical aspects allowed us to address the requirements so that the delivered INSPECTr framework complies with the EU standards in terms of data protection in the context of LEA operation work and of the GDPR during the development phase.

Measures were taken to anticipate potential issues and, when unexpected issues were identified, to mitigate them efficiently and support the safe continuation of the project.

Technical features were developed to safeguard access to the injected data and its processing, especially for the more advanced features involving AI technologies and evidence discovery.

The project engaged strong oversight from the outset, resulting in a system developed with ethics and privacy by design in mind throughout. The ultimate goal was to produce a platform that enhanced LEA abilities to provide a safer Europe for its citizens, while ensuring that the technology was compliant with Europe's strong human rights mandates and, therefore, acceptable to civil society.