



Lessen Data Access and Governance Obstacles

www.lago-europe.eu



ABOUT LAGO

LAGO, Lessen Data Access and Governance Obstacles, is a European Union-funded security-focused initiative with 24 partners from 14 European countries, including seven law enforcement agencies (LEAs).

LAGO will build an evidence-based and validated multi-actor reference architecture for a trustworthy EU FCT (fight against crime and terrorism) Research Data Ecosystem (RDE) to solve the FCT data issue. FCT-related data can be co-created, shared, and secured in the RDE.

For more information, visit <https://lago-europe.eu>.

LAGO NEWSLETTER

Issue 2, February 2024

TABLE OF CONTENTS

The Project Coordinator's Message • P. 2

Work Package Updates • P. 3-5

Past Events • P. 6-7

Future Events • P. 8

Publications • P. 8

Innovations Spotlight • P. 9



THE PROJECT COORDINATOR'S MESSAGE

ERNESTO LA MATTINA, ENG

Welcome to the second edition of the LAGO newsletter! As we share our progress and look towards future endeavours, it is evident that our collaborative efforts have significantly advanced our mission to establish a trusted EU Research Data Ecosystem (RDE) for Fighting Crime and Terrorism (FCT).

The challenges we have tackled, particularly around data privacy and the intricacies of legal and ethical compliance, have been substantial and highly rewarding.

The collaboration among small and medium-sized enterprises (SMEs), academic institutions, and law enforcement agencies (LEAs) has been instrumental in driving the creation of innovative solutions applicable to real-world scenarios.

This newsletter showcases the dedication and progress across various work packages, highlighting how each contributes to the overarching goals of the LAGO project.

Our work packages have shown encouraging advancements, and we are steadily moving towards our objectives. It's important to acknowledge the integral role each partner plays in this collaborative venture. Your insights and efforts are deeply valued and play a key role in driving us forward.

Looking ahead, our focus is on refining our roadmap for the implementation of the EU FCT RDE. This roadmap is essential for translating our research into actionable strategies and practices that adhere to European data protection and ethical standards.



Your ongoing engagement and feedback are vital to the continuous development and success of LAGO. Your perspectives help in fine-tuning our approach and ensuring that we stay on course.

Furthermore, I would like to highlight the presence of the LAGO community on [Zenodo](#). This platform plays an important role in making our collective knowledge and findings accessible, facilitating wider dissemination and engagement with our work.

We invite you to stay engaged through our updates, participate in upcoming events, and reach out via email or social media. Your active involvement is fundamental to the success of LAGO.

Thank you for your dedication and enthusiasm as we embark on another year of impactful work together.



WORK PACKAGE UPDATES

LAGO continues to harness the collective expertise of a diverse consortium, driving progress in the field of data management and governance within the security domain. Our collaborative force of European LEAs, SMEs, and academic partners remains dedicated to leading the charge towards innovative and feasibly applicable solutions.

In the following section, we present updates from our work packages (WP), demonstrating our sustained commitment to develop solutions that are not only innovative but also pragmatically viable in the realm of security.

WP2 ETHICAL, SOCIETAL AND LEGAL SCREENING, GUIDANCE AND REVIEW

WP2 has consistently advanced in its efforts to create an integrated ethical, societal, and legal framework. A significant achievement has been the completion of the deliverable on “Societal, Ethics, Legal, and Privacy (SELP) aspects and requirements for EU training and testing data spaces for FCT research”. Led by KU Leuven (KUL) and reviewed by the Centre for Security Studies (KEMEA), the Cybercrime Research Institute (CRI), the Estonian Police and Border Guard Board (PPA), and the Centre of Excellence in Terrorism, Resilience, Intelligence, and Organised Crime Research (CENTRIC), this deliverable evaluates the SELP requirements for EU FCT research data spaces. It encompasses tasks including the “Assessment of ethical and legal aspects for EU FCT research data spaces”, and the “Development of key SELP parameters and requirements for EU FCT research data spaces”.

Moving forward, WP2 will shift the focus towards providing policy recommendations that will guide legal data usage. Concurrently, the development of a compliance self-assessment tool is underway.

WP3 FRAMEWORK EU FCT TRUSTED RESEARCH DATA ECOSYSTEM (RDE)

Following exploration of the FCT landscape in research data access, collaborative efforts have led to significant advancements in WP3.

Partners have successfully identified use cases and requirements for the RDE, focusing on data access, sharing procedures, types, security measures, and anticipated benefits.

Eight key use cases have been outlined, encompassing areas like illicit arms trafficking, biometric data exchange, public space protection, deep fake detection, edge computing for public safety, counter-terrorism, and illicit goods trafficking. These also include providing data for research and training purposes.

From these use cases, a tiered approach to requirement elicitation has been developed. This ranges from business requirements aligned with project goals to stakeholder needs and, finally, solution requirements. These latter requirements are more technical, addressing both functional and non-functional aspects.

Progress has been made in the initial design of the Reference Model and Reference Architecture.

The Reference Model offers a high-level conceptualisation, detailing actors, processes, standards, and technical components within the RDE. The Reference Architecture then translates these concepts into a coherent technical solution. This approach has enhanced stakeholder and technician understanding, fostering efficient idea exchange and goal convergence.

A decentralised data repository model has been adopted, covering essential functionalities: onboarding, data creation, publishing, request handling, sharing, and model lifecycle management.

The next few months will see the finalisation of the Reference Architecture and the development of the Reference Implementation, which will undergo testing and evaluation in upcoming demonstration rounds.

WP4 RESEARCH DATA CREATION AND PROVISION

WP4 partners have made substantial progress in enhancing data management and security across the LAGO ecosystem. Our efforts have led to the expansion of data discovery and acquisition capabilities, including the introduction of a Twitter crawler and improved integration for searching Dark Web content. Improvements to our standalone REST API have boosted the efficiency of multimedia content extraction and the retrieval of images in various formats.

In the area of annotation and labelling, our frameworks have been developed through the integration of advanced active learning strategies designed for more effective data sample selection and manual annotation processes. We are also advancing in the development of tools like video summarisation and synthetic image generation, with a focus on enhancing their usability and integration into the LAGO environment.



A key area of progress is in privacy preservation, where we are refining anonymisation tools to ensure General Data Protection Regulation (GDPR) compliance, particularly relevant for anonymising licence plates and full-body images.

We are dedicated to improving the robustness and security of our datasets by enhancing the watermarking technique.

Despite facing challenges in tool integration within WP4, adopting a collaborative approach has facilitated smoother functionality testing among teams. Overall, WP4 is making significant contributions towards improving the efficiency and security of data handling within the LAGO ecosystem in the FCT domain, staying aligned with our project goals.

WP5 RESEARCH DATA USAGE

WP5 has already developed and demonstrated several prototypes as stand-alone tools that contribute to lifting the barriers to access to FCT research data. Among these, a data quality assessment prototype has been introduced to assist with the verification of the data used in training and testing. This prototype is based on the development and integration of novel indicators that enhance the reliability of the data.

In close collaboration with LEAs and data providers, WP5 has developed a risk assessment methodology prototype. This prototype is designed to measure, evaluate, and help mitigate the risks associated with granting access to FCT research data. Additionally, a blockchain component has been implemented and demonstrated, establishing a foundation for trust. Furthermore, a containerised federated learning framework has been developed, enabling the training of AI and ML models while maintaining secure and private access to data from LEAs.

WP5 is now focusing on the delivery of a sandbox service. This will allow security stakeholders to download and test AI and ML models on their premises. The current phase is dedicated to refining these prototypes towards their final versions and validating them in close collaboration with the end-users within the LAGO consortium.

In the next period, WP5 expects to increase its dissemination activities. This includes the submission of publications focusing on the novel elements developed within WP5.

WP6 RESEARCH DATASET GOVERNANCE

WP6 is focused on creating a governance framework to facilitate secure and efficient data exchange within the FCT research ecosystem. The initial concept and architecture have been outlined, leading to the development of a preliminary proof of concept (POC) for gathering feedback, particularly through participant engagement.

This POC showcases a structured approach to data exchange using metadata, where participants use verifiable credentials for the representation of identity and capabilities, adhering to the World Wide Web Consortium (W3C) guidelines. This iteration of the POC includes modules for credential issuance and a federated catalogue for listing and description of shared data.

The ISSUER(s) module is tasked with issuing Verifiable Credentials (VC) to each node within the RDE, including generic nodes. Future enhancements will integrate functionalities for organisational nodes, enabling each node to manage its VCs, selectively share them upon request, and authenticate the VC independently of the issuer's services. Moreover, the federated catalogue, which currently maintains a registry of nodes, is set to expand, aiming to catalogue comprehensively all data that members intend to share.

As part of our efforts towards initial validation, we have introduced a licensing tool into the data submission workflow and are currently assessing suitable data models and vocabularies for further development.

WP7 VALIDATION AND DEMONSTRATION IN REAL SCENARIOS

WP7 is dedicated to conducting practical demonstrations to test the viability and efficacy of LAGO's concepts, architecture, and solutions in real-life scenarios. The primary goals of this WP are threefold: to establish demonstration scenarios, carry out demonstration exercises, and collect valuable feedback from stakeholders. This feedback is essential for assessing the potential of the LAGO framework and implementing necessary improvements.

In the project's initial stages, particularly during Task 7.1, "Demonstration scenarios definition," our focus was on creating realistic scenarios based on use cases previously identified.



These scenarios are based on previously identified use cases and serve as a means to demonstrate LAGO's ability to facilitate data and model exchanges. They further aim to validate the entire architecture, tools, and services within the LAGO ecosystem.

Our demonstration activities engage a wide range of stakeholders, including researchers, LEAs, industry experts, and other relevant practitioners. The scenarios align with other projects in the field of FCT research and are centred on everyday situations requiring data sharing for training and testing purposes. This approach allows for the effective testing and development of AI solutions.

Over the course of three evaluation rounds, we will implement these scenarios to gather insights on usability and identify errors. The feedback and data collected through these exercises will help us further refine and enhance the LAGO ecosystem.



WP8 BOOSTING THE CREATION OF EU DATA ECOSYSTEM FOR FCT RESEARCH

Since the beginning of the LAGO project, WP8 has provided essential dissemination tools, such as LAGO-branded materials and comprehensive guidelines, to effectively share our progress and findings. Our active participation in conferences, seminars, and CERIS events, along with consistent publications, has bolstered our presence in the industry. Furthermore, our engagement through our website and social media channels enhanced our digital outreach.

WP8 has been instrumental in fostering community connections, evidenced by the establishment of collaborative agreements with projects and initiatives such as STARLIGHT, GRACE, Aligner, EACTDA, and ECTEG. Training activities have also been a focal point, starting with a questionnaire designed to tailor our training platform to the needs of our stakeholders.

Moving forward, WP8 remains committed to enhancing the dissemination of LAGO's contributions, with the aim of fostering a deeper understanding and acceptance within the community. Our efforts over the past year have laid a solid foundation for outreach and visibility, positioning LAGO as a prominent entity in the EU data ecosystem and the field of FCT research.

WP9 ETHICS REQUIREMENTS

WP 9 has achieved a significant milestone by submitting all four of its deliverables on schedule. Among these, the completion of the deliverable 9.3 stands out. This deliverable, which addresses Requirement No. 3, is especially important as it focuses on the development, deployment, and use of AI technologies within the LAGO framework.

The work on this deliverable began with a detailed analysis of the relevant legal frameworks, aiming to deepen the understanding of the deliverable's scope. It presents key findings from the Ethical Guidelines for Trustworthy AI and delves into the Ethics by Design and Ethics of Use Approaches for AI deployment. Additionally, it provides an in-depth look at the ALTAI (Assessment List for Trustworthy Artificial Intelligence) and its outcomes. Most importantly, it outlines comprehensive strategies for mitigating potential biases, discrimination, and stigmatisation in the input data, with a dual purpose: to meet the requirements of Ethics Requirement No. 3 and to guide the LAGO partners in the adoption of appropriate AI-based solutions.

In addition of these achievements, the LAGO consortium has established an External Ethics Board. The board which includes experts Luca Bolognini, Simone Casiraghi, and Denitsa Kuzhuharova, convened for the first time on November 20, 2023. During this meeting, they reviewed the primary deliverables of the LAGO project and have provided a preliminary report that offers a series of recommendations and suggestions. This underscores the consortium's commitment to ethical considerations and the responsible implementation of AI technologies.



PAST EVENTS

LAGO WAS PRESENTED AT RISE-SD IN RHODES

LAGO was presented at the Research and Innovation Symposium for European Security and Defence (RISE-SD) in Rhodes, Greece, from May 29–31, 2023. This was a three-day EU event on disaster response, crisis management, infrastructure security, and defence research, attended by EU officials, government representatives, researchers, industry experts, and field practitioners.



PROJECTS TO POLICY SEMINAR: BRIDGING THE GAP BETWEEN RESEARCH AND POLICY IN SECURITY

Ernesto La Mattina, the LAGO Project Coordinator, attended the Projects to Policy Seminar in Brussels on June 14–15, 2023. Organised by DG HOME.F2 (Innovation and Security Research) and REA.C2 (Secure Society), this seminar connected 36 security research projects from the 2021 Horizon Europe Cluster 3 "Civil Security for Society" call with Commission service representatives.



LAGO'S 3RD PLENARY MEETING IN THESSALONIKI

LAGO partner, the Centre for Research and Technology Hellas (CERTH), a prominent research institution in Greece, hosted the third LAGO Plenary Meeting in Thessaloniki on June 22, 2023. This meeting, occurring nearly eight months after the project's inception, served as a moment to review the progress made and strategise for the forthcoming phases in creating LAGO's reliable research data ecosystem.



PAST EVENTS

LAGO'S 4TH PLENARY MEETING IN LISBON

The LAGO project held its 4th Plenary Meeting in Lisbon on October 11, 2023, with Policia Judiciaria (PJ) as the host. This meeting served as a platform to review our project's efforts in solving data access and governance issues in the FCT area.

During the meeting, we discussed recent developments, examined the impact and solutions offered by LAGO, and focused on key objectives.



LAGO AT THE 2ND ANTI-FINTER POLICY SEMINAR IN BRUSSELS

The 2nd Anti-FinTer Policy Seminar in Brussels was held on October 26, 2023, and it concentrated on key topics such as equipping law enforcement with necessary tools, enhancing access to knowledge and data, and fostering collaboration among law enforcement agencies.

The event facilitated the exchange of ideas and highlighted the significance of collective efforts in navigating the intricacies of terrorist financing.



LAGO AT THE CERIS FCT/INFRA ANNUAL EVENT

LAGO joined the CERIS annual event on Fighting Crime and Terrorism/Resilient Infrastructure in Brussels on December 14–15, 2023, organised by DG HOME.

The event gathered experts for discussions on security research. Project Coordinator, Ernesto La Mattina and April Murray-Cantwell from partner CENTRIC represented LAGO, sharing insights on enhancing research data access. This participation underlines our commitment to tackling security challenges through shared knowledge and innovation.



FUTURE EVENTS

5TH PLENARY MEETING

The fifth plenary meeting is scheduled for the end of February 2024. This gathering will allow the project team members to review recent developments and any adjustments to the project's timeline and milestones.

It will also offer stakeholders an opportunity to provide their input and thoughts on the project's progress to date.



PUBLICATIONS

ARTICLE IN JOURNAL

Peppes, N., Alexakis, T., Daskalakis, E., Demestichas, K. & Adamopoulou, E., 2023. Malware Image Generation and Detection Method Using DCGANs and Transfer Learning. IEEE Access, 11, pp.1-1. Available at: <https://ieeexplore.ieee.org/document/10264089> [Accessed September 2023].

CONFERENCE PROCEEDINGS

Peppes, N., Alexakis, T., Daskalakis, E., Adamopoulou, E., & Demestichas, K. (2023). A Generative Adversarial Network (GAN) Solution for Synthetically Generated Botnet Attacks Data Samples. In Security and Defense 2023 Conference (pp. 216-219), May 29-31, Rhodes, Greece.

Ammar, H., Loesch, A., Vannier, C. & Audigier, R., 2023. Can Human Attribute Segmentation be More Robust to Operational Contexts Without New Labels? In: 2023 IEEE International Conference on Image Processing (ICIP). [Online] Available at: <https://ieeexplore.ieee.org/document/10222300> [Accessed 12 October 2023].

Psaltis, A., Kastellos, A., Patrikakis, C.Z. & Daras, P., 2023. FedLID: Self-Supervised Federated Learning for Leveraging Limited Image Data. In: ICCV2023. [Online] Available at: https://openaccess.thecvf.com/content/ICCV2023W/LIMIT/html/Psaltis_FedLID_Self-Supervised_Federated_Learning_for_Leveraging_Limited_Image_Data_ICVW_2023_paper.html [Accessed October 2023].

Psaltis, A., Chatzikonstantinou, C., Patrikakis, C.Z. & Daras, P., 2023. FedRCIL: Federated Knowledge Distillation for Representation based Contrastive Incremental Learning. In: ICCV2023. [Online] Available at: https://openaccess.thecvf.com/content/ICCV2023W/VCL/html/Psaltis_FedRCIL_Federated_Knowledge_Distillation_for_Representation_based_Contrastive_Incremental_Learning_ICCVW_2023_paper.html [Accessed October 2023].

CHAPTER IN A BOOK

Peppes, N., Alexakis, T., Daskalakis, E., Adamopoulou, E., & Demestichas, K. (Pending publication). A Generative Adversarial Network (GAN) Solution for Synthetically Generated Botnet Attacks Data Samples. In B. Akhgar (Ed.), Security Informatics and Law Enforcement (Series Editor: Babak Akhgar). Springer.

GREEN PAPER

CENTRIC. (2023). Challenges and Recommendations for a Research Data Ecosystem to Support Innovations in the FCT Domain (Curated by B. Akhgar & P. Bayerl).



MALWARE IMAGE GENERATION AND DETECTION METHOD USING DCGANS AND TRANSFER LEARNING

BY NIKOLAOS PEPPE, THEODOROS ALEXAKIS, EMMANOUIL DASKALAKIS, KONSTANTINOS DEMESTICHAS AND EVGENIA ADAMOPOULOU
INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS (ICCS), GREECE

Cybersecurity has become a paramount concern in today's digital age, impacting various activities. A pressing challenge arises with the emergence of a new type of cyber threat: the potential embedding of malware into digital images, allowing it to seamlessly bypass conventional scanners and compromise the security of both service providers and end-users.

UTILISING DCGANS FOR MALWARE IMAGE CREATION

In response to this, the research introduces a holistic methodology designed to generate realistic-looking malware-based images and develop an effective malware detector. This innovative approach leverages Deep Convolutional Generative Adversarial Networks (DCGANs) to create two distinct datasets: Expanded Malware Images (EMI), containing suspicious malware images, and Fashion Adversarial Samples (FAS), comprising adversarial samples of fashion products.

The generated datasets serve as the foundation for training two different Convolutional Neural Network (CNN) models, each following a distinct training approach. The first CNN, c-CCN, adopts a conventional training methodology, while the second CNN, TL-CCN, employs transfer learning by capitalising on the knowledge embedded in ResNet50. Notably, the research demonstrates that the generation of malware images and adversarial samples reaches stability after 3000 iterations, resulting in the production of highly realistic images.

SUPERIOR PERFORMANCE OF THE TL-CCN MODEL

The significance of the study becomes apparent as it unfolds the effectiveness of the TL-CCN model, trained with a subset of the adversarial samples.

This model outperforms other malware detectors, showcasing superior results in terms of high validation accuracy and minimal validation loss.

Furthermore, the study underscores the critical importance of quality assessment in the generated data samples through DCGANs. Rigorous evaluation ensures that the synthetic malware images and adversarial samples not only appear realistic to the human eye but also possess the various characteristics required for effective training and evaluation of malware detection models.

This meticulous quality assessment stands as a cornerstone for enhancing the reliability and robustness of the proposed methodology against sophisticated cyber threats.

CONTRIBUTION TO CYBERSECURITY: COMBATING SOPHISTICATED THREATS

Thus, this research work focuses on crucial elements such as malware generation, generative adversarial networks (GANs), transfer learning, convolutional neural networks (CNN), and cybersecurity.

By addressing these key terms, the study contributes valuable insights to the field of cybersecurity. It does not only acknowledge the challenges posed by malware embedded in images but also introduces advanced techniques, including DCGANs and transfer learning, to combat these emerging threats.

The application of such methodologies proves instrumental in both generating realistic malware images and enhancing the accuracy of malware detection systems.

EVALUATING SYNTHETIC MALWARE FOR ENHANCED SECURITY

In conclusion, this study provides a comprehensive understanding of the contemporary cybersecurity landscape, emphasising the need for innovative approaches to counter evolving cyber threats. The integration of DCGANs and transfer learning emerges as a potent solution, showcasing its potential to consolidate malware detection capabilities in the face of image-based cyber threats. Crucially, the study emphasizes the particular quality assessment conducted on the generated data samples through DCGANs. This rigorous evaluation ensures that the synthetic malware images and adversarial samples not only achieve a high degree of realism but also exhibit the various characteristics essential for robust training and evaluation of malware detection models. As cybersecurity methodologies continue to evolve, adapting to the dynamic nature of digital threats in the modern era, this research makes a significant contribution by advancing our understanding and securing the defences against sophisticated cyber adversaries.



KEY FACTS

Coordinator: Engineering Ingegneria Informatica S.p.A. (Italy)

Start Date: November 2022

Duration: 24 months

Consortium: 24 partners from 14 countries, including 7 LEAs

Topic: HORIZON-CL3-2021-FCT-01-04, Improved access to fighting crime and terrorism research data

Type: HORIZON Innovation Actions


Total cost: €7.38m


CONSORTIUM


			
			
			
			
			
			



FOLLOW US

 lago-europe.eu

 info@lago-europe.eu

 [@lago-europe](https://www.linkedin.com/company/lago-europe)

 [@lago_europe](https://twitter.com/lago_europe)

