Lessen Data Access and Governance Obstacles

www.lago-europe.eu

LAGO NEWSLETTER
Issue 3, November 2024

ABOUT LAGO

LAGO, Lessen Data Access and Governance Obstacles, is a European Union-funded security-focused initiative with 25 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/**.

THE PROJECT COORDINATOR'S MESSAGE

ERNESTO LA MATTINA, ENG

Welcome to the third edition of the LAGO newsletter! We are pleased to share the steady progress we have made in shaping a secure and reliable EU RDE for FCT. As we move closer to the project's conclusion, our mission remains focused on improving data accessibility and governance for law enforcement and research communities.

Our collaborative work with law enforcement agencies, researchers, industry, and policymakers continues to support the development of a sustainable framework for secure and ethical data sharing.

These partnerships are essential to fostering a system that enables cross-border cooperation while protecting sensitive data.

In this edition, we highlight key achievements from each work package. For example, work package 2 has developed a self-assessment compliance toolkit, helping users navigate legal and ethical complexities within an evolving regulatory landscape. Work package 3 has advanced the Reference Architecture for the RDE, implementing its hybrid approach across partner sites and successfully completing two rounds of demonstrations. Work packages 4 and 5 have focused on enhancing data management and anonymisation techniques, incorporating tools like blockchain, and federated learning frameworks.

At the end of June, work package 8 organised a policy event in Brussels, to present the LAGO Roadmap Vision and gather feedback from subject-matter experts, an event that provided a dynamic policy discourse forum, helping to validate the LAGO RDE roadmap. Notably, in September, the project was presented at the Law Enforcement Working Party (LEWP) in

Brussels. This session underlined LAGO's value in advancing secure data-sharing practices, with the EU Commission stressing the importance of our work.

Our involvement in key events, such as Tools4LEAs, along with collaborations with partners like ECTEG and EACTDA, has allowed us to showcase LAGO's developments and strengthen our network of stakeholders.

The success of LAGO relies on the dedication and teamwork of our consortium, and I would like to thank everyone for their sustained efforts over the past two years. We are not just solving technical challenges but also laying the groundwork for a more transparent and effective approach to how sensitive data is handled in cross-border investigations. This includes maintaining the highest ethical standards and ensuring compliance with new regulations, such as the AI Act, which will shape future developments in this area.

We invite you to explore this edition for a closer look at our recent achievements and insights into the impactful work being done. Thank you for your continued support and collaboration.

WORK PACKAGE UPDATES

The following section provides updates from each work package (WP), updates that reflect the consortium's dedication to addressing real-world challenges with effective, research-backed solutions.

WP2 ETHICAL, SOCIETAL AND LEGAL SCREENING, GUIDANCE AND REVIEW

The FCT Research Data Ecosystem's compliance strategy is currently being developed. This is in response to the legal and ethical complexities encountered during the project. The fragmented regulatory landscape led to a phased approach, starting with assessments and the development of a compliance solution. Our research revealed challenges stemming from variations in national laws and the European Union's harmonisation efforts, with local legislation often taking precedence.

Attempts to implement a compliance-by-design infrastructure were set aside due to the complexities of multi-jurisdictional laws, the dynamic nature of legal frameworks, and software limitations in interpreting nuanced regulations. Consequently, a self-assessment compliance toolkit is now being developed. This toolkit, based on 51 structured questions, guides participants in navigating legal and ethical issues during dataset exchanges. It enables users to critically engage with regulations and adapt to the evolving legal landscape, promoting greater accountability and supporting data protection, privacy, and ethical standards, including upcoming AI regulations. Ultimately, the toolkit will foster informed decision-making while remaining flexible and adaptable to ongoing regulatory changes.

WP3 FRAMEWORK EU FCT TRUSTED RESEARCH DATA ECOSYSTEM (RDE)

WP3 has focused its efforts on developing the Reference Architecture and Reference Implementation for the LAGO RDE. The project received an extension of six months to address the cases of entities with limited lifespans, such as research projects, so that the data produced continues to remain available even after their end. To achieve that, the architecture has been upgraded from purely decentralised to hybrid decentralised-centralised, enabling the organisations to maintain control over their datasets while advertising metadata in a shared catalogue. This approach ensures that sensitive FCT data is only accessible to accredited organisations within the ecosystem, which are verified by a trusted authority. Once accredited, organisations can search for datasets in the catalogue and request access to relevant data providers. After agreeing on the contractual terms of use, data can be transferred directly from the provider to the requesting organisation through peer-to-peer connections, facilitated by dedicated software components, the Connectors, that implement the RDE standards and protocols.

Based on the above, the Reference Implementation of the RDE has been realised, consisting of multiple interacting components: the Catalogue, which collects metadata of all datasets available in the RDE; Connectors, which allow organisations to publish metadata on the Catalogue and exchange datasets; a distributed Ethereum-based ledger, that keeps track of all activities in the RDE; and the Issuer Connector, which allows a trusted authority to receive requests for joining the RDE and register new organisations in the ecosystem.

The Reference Implementation has been successfully deployed in a distributed way across the premises of three partners (Engineering Ingegneria Informatica s.p.a., INOV - INESC Inovação Instituto de Novas Tecnologias, and AIT Austrian Institute of Technology GmbH) and tested by all partners during two demonstration rounds. Moving forward, WP3 will focus on developing a centralised repository, catering to the needs of entities with limited lifespans, such as research projects, ensuring that their datasets remain accessible to the FCT research community even after their completion.

WP4 RESEARCH DATA CREATION AND PROVISION

Over the past seven months, WP4 has made significant progress in improving data management, labelling, and security within the LAGO ecosystem. This has been achieved through the implementation of novel active learning methodologies, which have successfully reduced the cost of labelling by dynamically adjusting batch sizes based on real-time feedback from models.

These advancements have been consolidated into a report, which documents the progress made thus far.

WP4 also encountered challenges, particularly in balancing the realism of data anonymisation with computational efficiency. Efforts have focused on refining these techniques to ensure that multimedia content, particularly images, remains visually realistic without overburdening system resources.

Feedback from pilot tests has been crucial in helping the team adapt the tools for real-world applications, and WP4 continues to make adjustments to ensure they meet the demands of the project.

Going forward, WP4 will concentrate on improving LAGO's data management capabilities, including sharing, anonymising, and annotating both textual and visual content, ensuring that all tools are optimised for use across the LAGO platform.

WP5 RESEARCH DATA USAGE

WP5 has made substantial progress on the development of data quality, risk assessment, blockchain, and federated learning tools. The final versions of the data quality and risk assessment tools have been delivered, following extensive validation tests and close collaboration with end users during pilot phases.

The data quality tool has integrated new indicators and analysis metrics, while the risk assessment tool has been enhanced with new examples, explanations, and mitigation strategies. Work on the blockchain component is advancing, with the integration into WP6's governance framework nearing completion. This will form the basis of the ledger system, ensuring the secure exchange of data across the ecosystem.

The federated learning framework has also progressed, offering a very simple and streamlined setup that allows privacy-conscious organisations secure access without exposing sensitive data.

The sandbox environment is being refined based on user feedback, improving the security and usability of testing models and tools in controlled conditions. In the coming months, WP5 will focus on finalising the development of all tools, continuing validation during pilot phases, and engaging in dissemination activities to ensure wide adoption.

WP6 RESEARCH DATASET GOVERNANCE

WP6 continues to work on the design and implementation of a governance model that enables secure, trusted data exchanges within the LAGO ecosystem. The federated data repository ensures that data providers retain full control over their datasets, deciding which data to make available and under what conditions, including licensing and usage policies.

The governance model includes modules for the secure identification of both data consumers and providers, based on W3C Verifiable Credentials issued by a trusted party. A catalogue is provided, allowing users to browse available datasets described using standard vocabularies, and support is available for defining terms of use, licenses, and, if required, signing contracts. Once terms are agreed upon, data is transferred in a peer-to-peer fashion, ensuring that transactions are transparent and trustworthy. All agreements and transfers are recorded on blockchain, adding an additional layer of security.

Participants maintain their datasets on their premises on dedicated servers, called Nodes, and access is managed through Connectors that house all governance modules. A prototype of the Connector is now available for testing and evaluation, including governance functionalities to ensure seamless interaction between Nodes.

WP7 VALIDATION AND DEMONSTRATION IN REAL SCENARIOS

In recent months, WP7 has focused on the testing and validation of the LAGO ecosystem through a series of demonstration sessions. Two events were held, one in Paris and another in Helsinki, where the tools and overall system were showcased. Following the Paris workshop, the first pilot was conducted, and feedback was gathered during a conciliation meeting in Helsinki using the Delphi method.

This meeting allowed technical partners and end users to address issues encountered during the first pilot phase. The second pilot, recently finalised, focused on validating LAGO as a comprehensive ecosystem for ethical, deontological, and legal data handling and exchange.

Particular attention has been paid to the seamless integration of tools and the main platform, including data creation, anonymisation, and sharing with the different components of the LAGO ecosystem. The results of this pilot will be analysed in preparation for a third and final pilot.

WP8 BOOSTING THE CREATION OF EU DATA ECOSYSTEM FOR FCT RESEARCH

WP8 has continued to support the development of the LAGO Research Data Ecosystem for FCT research through its dissemination, training efforts, and sustainability. Progress is being made on the sustainability and exploitation plan, and training materials are also under development to meet stakeholder needs.

We have actively fostered community connections through engagement with policymakers, researchers, and the Data Spaces community, with the aim of encouraging the adoption of the LAGO Roadmap. At the end of June, the Centre of Excellence in Terrorism, Resilience, Intelligence, and Organised Crime Research (CENTRIC) hosted a policy event in Brussels to showcase the LAGO Roadmap Vision and gather feedback from subject-matter experts. The event facilitated a dynamic policy discourse forum, employing the ROAMEF policy development cycle to collect expert feedback and validate the LAGO RDE roadmap.

In addition to participating in conferences, seminars, and key events, which provided valuable platforms to showcase project outputs, WP8 has maintained a robust digital presence through its website and social media channels. Looking ahead, WP8 will continue to enhance LAGO's visibility, ensuring that its work is well-integrated within the FCT research environment. Collaboration with other work packages and external partners will remain a key focus, as WP8 aims to build a strong and engaged community around the LAGO initiative.

WP9 ETHICS REQUIREMENTS

Over the past few months, significant progress has been made in WP9. The LAGO consortium has focused on addressing comments and incorporating feedback from the ethics board and reviewers into LAGO. This meticulous work ensures that our project aligns with the highest ethical standards and meets all necessary compliance requirements. Additionally, two training sessions were organised with the support of the Cybercrime Research Institute (CRI) and the Katholieke Universiteit Leuven (KU Leuven), covering the AI Act and GDPR. These sessions have significantly enhanced our team's understanding and compliance, equipping us with the knowledge to navigate these complex regulatory landscapes effectively.

With the project's six-month extension, the LAGO consortium now has the opportunity to schedule additional training sessions. These can either introduce new topics of interest or delve deeper into the AI Act and GDPR, depending on the needs and interests of our partners. This extension provides a valuable window to further strengthen our team's capabilities and ensure we are fully prepared for future challenges.

Furthermore, due to the project extension, the second Ethics Report, initially planned for November 2024, has been postponed. This delay allows us to ensure that all deliverables are thoroughly prepared and reviewed. The revised timeline will help us maintain the quality and integrity of our work, ensuring that we meet all project milestones effectively.

PAST EVENTS

LAGO'S 5TH PLENARY MEETING IN VIENNA

The LAGO project held its 5th Plenary Meeting in Vienna between February 29th and March 1st 2024, hosted by AIT Austrian Institute of Technology GmbH. This meeting provided an opportunity to assess the project's progress over the past year, address feedback from the first review, and set strategic goals for the next phases. Key discussions included updates on work packages, initial demonstrations, and upcoming initiatives, all aimed at advancing secure data sharing in the FCT domain.

LAGO AT THE TOOLS4LEAS EVENT IN LISBON

The LAGO team took part in the Tools4LEAs event in Lisbon from 12th to 14th March 2024. Co-hosted by the European Anti-Cybercrime Technology Development Association (EACTDA) and Polícia Judiciária, this event brought together EU law enforcement agencies, sister projects, and technology providers to focus on innovative tools for enhancing law enforcement capabilities.

LAGO AT TECNOSEC 2024 IN MADRID

On the 8th and 9th of May 2024, LAGO was represented by Vicomtech at TECNOSEC, Spain's top forum for security technologies. Our colleague, David Ríos, took part in a round table on Artificial Intelligence in security, highlighting LAGO's work on secure data exchange and governance. The event allowed LAGO to engage with key stakeholders and showcase its contributions to AI-driven security technologies and European collaboration.

LAGO 6TH PLENARY MEETING IN PARIS

The 6th Plenary Meeting took place on May 23rd, 2024, in Paris, hosted by the French Ministry of Interior (MININT). This in-person meeting allowed the team to review progress and plan the next steps for each work package. The discussions focused on refining strategies for upcoming workshops, enhancing dissemination efforts, and exploring future initiatives to strengthen the project's impact. Preliminary research results were also showcased, setting the stage for future developments.

LAGO'S FIRST WORKSHOP IN PARIS

The first LAGO workshop was held in Paris on May 24th, launching our initial demonstration campaign. Our technical partners showcased research outcomes from LAGO, STARLIGHT, and TRACY to investigators from the French Gendarmerie and EU LEA representatives.
Participants discussed the LAGO's Research Data Ecosystem, dataset creation, and governance framework. Two demo sessions provided hands-on insights into our innovations.

LAGO AT THE PROJECTS TO POLICY SEMINAR IN BRUSSELS

On 25th and 26th June 2024, LAGO was present at the Projects to Policy event organised by the European Research Executive Agency (REA) and the Directorate-General for Migration and Home Affairs (DG HOME) in Brussels.
The event provided a platform for direct interaction between project coordinators and policymakers, facilitating feedback on project activities and outputs, supporting evidence-based policymaking, and tracking project impacts.

LAGO'S 7TH PLENARY MEETING IN HELSINKY

On 5th of September 2024, the LAGO consortium met in Helsinki at the Ministry of Interior Finland (FIMOI) premises to review progress and plan the final steps of the project. Key discussions included tools, pilots, integrated scenarios, and the launch of the second demonstration campaign. A co-working session ensured alignment on tool usage, focusing on sustainability and planning for future workshops and dissemination activities.

LAGO'S 2ND WORKSHOP IN HELSINKY

LAGO held its second official workshop in Helsinki on 6th September 2024, focusing on project results and synergies with sister projects ENACT, MULTIRATE, and FALCON. The event featured presentations on LAGO's Research Data Ecosystem and cross-project use cases.
A hands-on session allowed participants to test the ecosystem, with special attention to the legal, ethical, and technical aspects of data sharing. Feedback was collected from all attendees to enhance the LAGO ecosystem.

LAGO AT THE LAW ENFORCEMENT WORKING PARTY OF THE CONSILIUM OF THE EUROPEAN UNION

On 17th September 2024, colleagues from the Italian National Police, invited by the Hungarian Presidency of the EU, had the opportunity to present the LAGO project to the 27 delegations representing the European LEAs, Europol, and the EU Commission.

Our concise presentation was well-received by the Presidency and the EU Commission, whose representatives underlined the importance of the LAGO project as a pioneering approach addressing the long-standing issue of dataset scarcity—a challenge that has become increasingly crucial with the rise of AI tools and the AI Act.

The EU Commission advised delegates to consider the LAGO project as a good example of addressing the dataset challenges and expressed hope that the LAGO approach could serve as a model for future collaborations.

LAGO AT THE CERIS ANNUAL EVENT ON FCT IN BRUSSELS

LAGO attended the CERIS annual event on Fighting Crime and Terrorism (FCT) in Brussels on September 24th and 25th, 2024. Organised by DG HOME, the event brought together experts to discuss online harms and security research. Prof. Dr. Marco Gerke from partner Cybercrime Research Institute (CRI) joined a panel on the "Implications of the AI Act for research efforts in fighting online crimes." Our participation highlights the consortium's dedication to tackling security challenges in the digital age through cooperation and innovative approaches.

LAGO AT RISE-SD 2024 IN CHALKIDIKI

The LAGO project was presented at the Research and Innovation Symposium for European Security and Defence (RISE-SD) in Chalkidiki, Greece, from October 16–17, 2024. Project Coordinator Ernesto La Mattina of Engineering Ingegneria Informatica s.p.a. (ENG) discussed the project's approach to building a trusted European FCT Research Data Ecosystem. By establishing a comprehensive reference architecture and ensuring high data quality, we aim to provide the foundation for trustworthy AI-driven research while complying with EU standards.

LAGO AT THE TOOLS4LEAS EVENT IN MALLORCA

On October 22nd, 2024, we attended the EACTDA/Tools4LEAs event in Mallorca. During our session, we presented the LAGO Research Data Ecosystem (RDE), showcasing its ability to enable secure data sharing between LEAs. A live demo was provided, with our colleagues, Dr. Jorge García Castaño from Vicomtech and Nikolaos Peppes from the National Technical University of Athens—Institute of Communications and Computers Systems (ICCS-NTUA), walking the audience through the process of sharing datasets, both from the data provider and consumer perspectives.

FUTURE EVENTS

LAGO'S FINAL WORKSHOP

The third and final LAGO workshop will build on the successes of the previous events, offering a comprehensive overview of the project's key achievements. Participants will have the opportunity to explore the latest advancements in LAGO's Research Data Ecosystem, with an emphasis on lessons learned from the previous hands-on sessions in Paris and Helsinki. Special focus will be given to real-world applications and cross-project collaborations, ensuring that legal, ethical, and technical challenges are addressed.

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.

CONFERENCE PROCEEDINGS
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).

Aramburu, M., Redó, D. and García-Castaño, J., 2024. *Securing multimedia-based personal data: towards a methodology for automated anonymization risk assessment seeking GDPR compliance.* In: SPIE Security + Defence Conference.

Bideaux, M., Phé, A., Chaouch, M., Luvison, B. and Pham, Q.C., 2024. *3D-COCO: Extension of MS-COCO Dataset for Scene Understanding and 3D Reconstruction.* In: 2024 IEEE International Conference on Image Processing (ICIP). IEEE.

Baloukas, C., Papadopoulos, L., Demestichas, K., Weissenfeld, A., Schlarb, S., Aramburu, M., Redó, D., García, J., Gaines, S., Marquenie, T., Eren, E. and Peter, I., 2024. *A Risk Assessment and Legal Compliance Framework for Supporting Personal Data Sharing with Privacy Preservation for Scientific Research.* In: The International Conference on Availability, Reliability and Security (ARES 2024) - TRUSTBUS Workshop. ACM.

Karl, R., Mazzonello, V., Pacheco, B., Silveira, J., Schlarb, S., Siska, V. and Weissenfeld, A., 2024. *Building a Data Space for the Public Security Domain using Smart Contracts.* In: The Fourth Intelligent Cybersecurity Conference (ICSC2024). IEEE.

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.

Psaltis, A., Chatzikonstantinou, C., Patrikakis, C.Z. & Daras, P., 2023. *FedRCIL: Federated Knowledge Distillation for Representation based Contrastive Incremental Learning.* In: ICCV2023.

Psaltis, A., Kastellos, A., Patrikakis, C.Z. & Daras, P., 2023. *FedLID: Self-Supervised Federated Learning for Leveraging Limited Image Data.* In: ICCV2023.

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).

LAGO TRAINING: EMPOWERING SECURE DATA SHARING IN FCT

BY EVANGELOS MICHOS
IANUS TECHNOLOGIES, CYPRUS

The LAGO project has been making great progress with its series of training sessions, designed to help LEAs, researchers, academics and security practitioners get up to speed with secure data sharing in the FCT area.

These sessions have given participants the tools, practical and theoretical knowledge, and also the guidance they need to use the LAGO Research Data Ecosystem (RDE) effectively. Over four training sessions to date, attendees learned about the technologies, legal frameworks and practical applications at the heart of LAGO's secure data-sharing infrastructure.

These training sessions have played a key role in helping LAGO's stakeholders understand and use the tools and technologies. With these skills, participants are now better equipped to overcome the challenges of secure data exchange and collaborate more effectively in the FCT area.

SESSION 1: FOUNDATIONS OF THE DATA ECOSYSTEM FOR SECURE DATA SHARING

The first training session was hosted online by IANUS Technologies (IANUS) on May 27-28, 2024. This session laid the groundwork for the whole series, helping participants understand LAGO's core goals and how its tools fit into secure data sharing for FCT research.

Presentations focused on LAGO's Reference Architecture, which supports secure data sharing while ensuring that data sovereignty, security, and quality are maintained. Participants also learnt about the legal and ethical standards for handling data across borders, especially with GDPR regulations in mind.
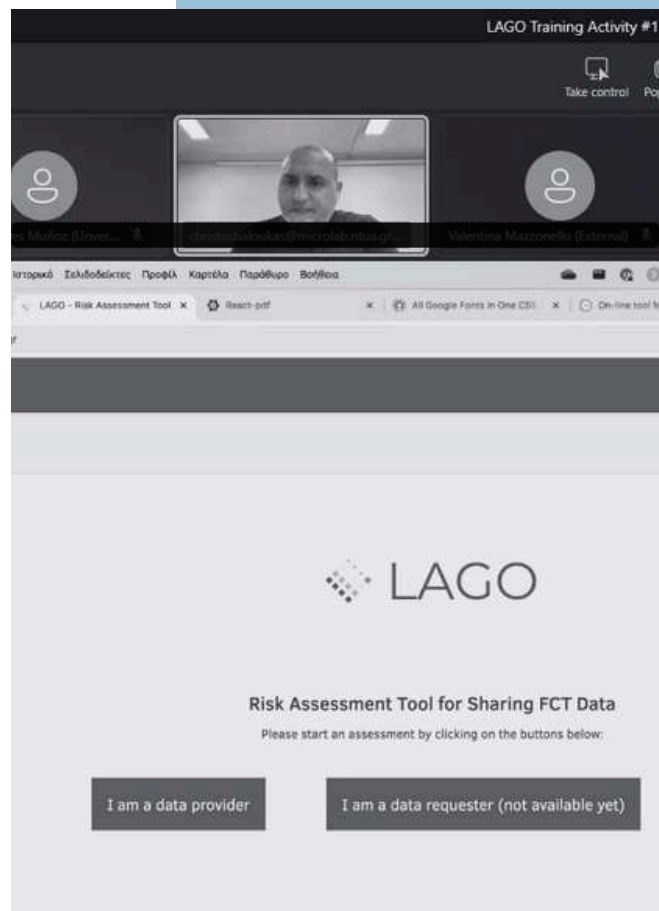
Additionally, they were introduced to the LAGO training platform developed by IANUS which provided both live and self-paced learning opportunities. The session featured live demos and videos to help participants get comfortable with the tools.

SESSION 2: UNDERSTANDING THE AI ACT

On June 6th, 2024, a second training session hosted by the Cybercrime Research Institute (CRI) covered the new AI Act and its impact on AI systems used in FCT research.

Participants learnt about the key features of the AI Act, including the four risk categories for AI systems—ranging from minimal risk to unacceptable risk—and what it means for their compliance responsibilities.

The session paid particular attention to high-risk AI systems, such as biometric identification and emotion recognition, and the mandatory legal requirements for these systems. Participants had a clear understanding of the practical steps they needed to take to ensure their AI technologies aligned with the law.

SESSION 3: LEGAL FRAMEWORK AND ETHICAL IMPLICATIONS OF PERSONAL DATA PROCESSING

The third session, held in June 2024 and hosted by the Katholieke Universiteit Leuven (KU Leuven), focused on the legal framework and the ethical implications of personal data processing.

The training aimed to clarify key principles of personal data handling, with an emphasis on GDPR and the Law Enforcement Directive. Participants explored the distinctions between anonymisation and pseudonymisation and how they can be applied to reduce re-identification risks.

They also learnt about the ethical challenges related to data processing, such as using publicly accessible data for research or handling sensitive data from the dark web.

By the end, attendees had a solid understanding of how to navigate the complex legal and ethical landscape of data handling in FCT research.

SESSION 4: ADVANCED TOOLS AND STRATEGIES FOR SECURE DATA SHARING

The fourth training session, hosted by IANUS in September 2024, expanded on the existing foundations and focused on secure data acquisition, management, and sharing within the LAGO RDE.

Participants learnt how to securely exchange datasets while adhering to data governance principles.

The session also introduced methods for automating the transcription of audio data, annotating and managing text and PDF documents, and using decentralised learning frameworks to protect data privacy during machine learning.

Additional topics included creating licence agreements for different types of data, synthesising data to protect privacy, and embedding security measures within datasets to detect misuse.
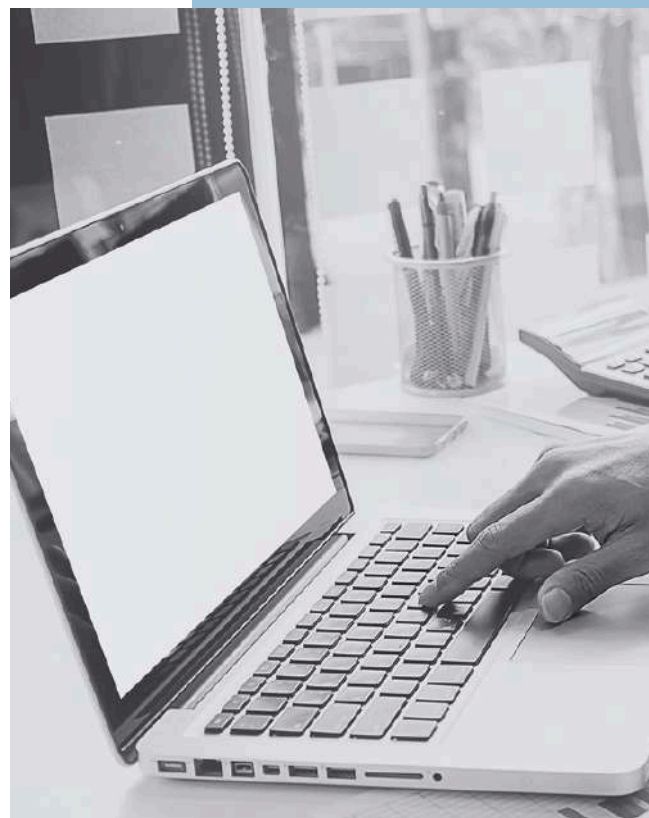
LOOKING AHEAD: UPCOMING TRAINING SESSIONS

As the LAGO project progresses, additional training sessions are planned to cover more advanced topics and further refine participants' skills in secure data management and data-sharing practices.

The previous training sessions covered both theoretical and practical aspects to set the stage for the upcoming sessions, which will focus on deepening the practical application of LAGO's RDE and its multiple use cases defined, ensuring participants are not only proficient but are also equipped to tackle emerging challenges in the FCT domain.

The ongoing training efforts play an important role in creating a trusted European RDE and by continuously engaging stakeholders through targeted training, LAGO will ensure that the latest technologies and frameworks are effectively integrated into real-world practices.

3D-COCO: EXTENSION OF MS-COCO DATASET FOR SCENE UNDERSTANDING AND 3D RECONSTRUCTION

BY ROMARIC AUDIGIER, ALICE PHE
CEA-LIST, FRANCE

Researchers from the Laboratory for Integration of Systems and Technology at the Commission for Atomic Energy and Alternative Energies (CEA-List), as part of the LAGO project, have introduced an innovative method for creating datasets that combine 2D images and 3D CAD models. This approach enables the training of new object detectors in a 'zero-shot' manner—meaning that even when no images are available for the objects of interest, CAD models can be used to train the detectors. By utilising this methodology, the need for manually acquiring and labelling images is significantly reduced.

A potential application for LEAs is the ability to inject CAD models of objects they are searching for—such as weapons or specific vehicles—into the system to develop specialised detectors.

These detectors can then retrieve the occurrence of such objects from large datasets or video streams, enhancing capabilities for monitoring and surveillance. The work presented includes the scientific methodology, a dataset, as well as experiments and results using public data.

INTRODUCING 3D-COCO: A COMPREHENSIVE DATASET FOR 2D-3D VISION TASKS

Our colleagues introduced 3D-COCO, a dataset that provides 2D real images, 3D CAD models, and their associated 2D-3D alignment.
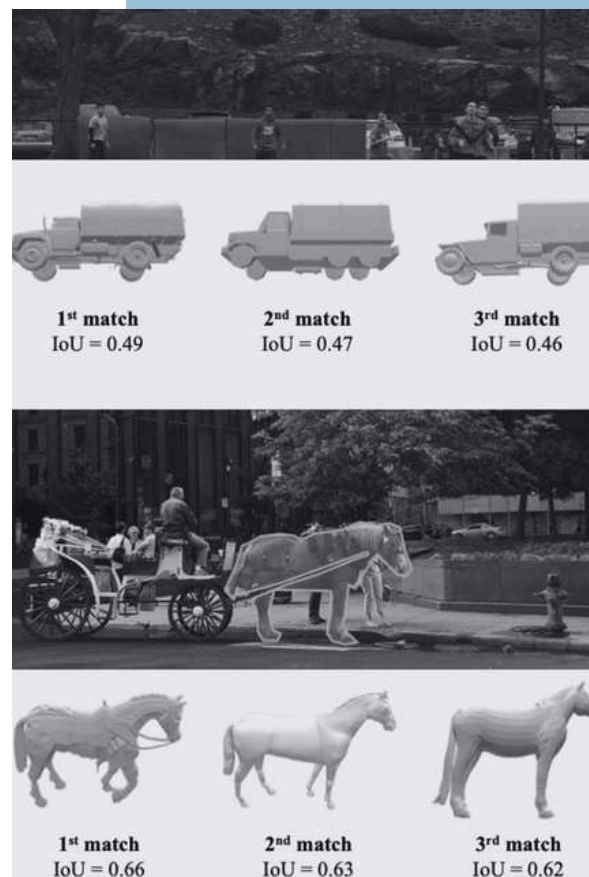
This dataset was curated to support computer vision tasks such as 3D reconstruction and image detection, configurable with textual, 2D image, and 3D CAD model queries.

While natural language and 2D images have shown remarkable proficiency in many tasks—particularly detection through the expansion of training data—3D-based vision remains underexplored despite its high information density. 3D-COCO builds on COCO, a common benchmark for object detection, and includes a methodology to extend other datasets where 3D models are available.

3D MODEL COLLECTION AND CURATION

To build the CAD model collection, the team sourced models relevant to 80 semantic classes from MS-COCO, primarily from ShapeNet and Objaverse, due to their alignment with computer vision tasks like 3D reconstruction and generation. This large-scale set includes 53,000 unique, high-quality 3D shapes, with 28,000 selected to annotate 3D-COCO.

2D-3D ALIGNMENT AND ANNOTATION

The annotation process was an automatic, class-driven process based on simple and robust geometry. Each 2D instance in the dataset was matched with its most similar 3D CAD model using an Intersection over Union (IoU) metric, which compared untextured renders of 3D models from multiple viewpoints with 2D instance silhouette masks and segmentations.

The highest-scoring CAD model was matched to each annotation, resulting in 860,001 training annotations and 36,781 validation annotations in 3D-COCO.

APPLICATIONS AND EXPERIMENTAL VALIDATION

With this alignment, 3D-COCO enables training on methods that leverage 3D data, including multimodal object detectors, synthetic multiview 3D reconstruction networks, real single-view 3D reconstruction networks, and in-the-wild pose estimation methods.

In experiments, our researchers trained an object detector using 3D models, allowing them to infer on novel classes in a 'zero-shot' manner—utilising CAD models as an external source without any images of the target objects.

Unlike text prompts, which often require additional prompt engineering, 3D CAD models provide a more consistent source for this application.

All 3D CAD models in 3D-COCO are licenced in a non-restrictive manner, and the augmentation of MS-COCO maintains privacy and ethical standards of the original dataset.

This research introduces a novel, publicly available dataset and a robust methodology for generating similar datasets. We hope 3D-COCO will encourage further exploration of 3D-related topics.
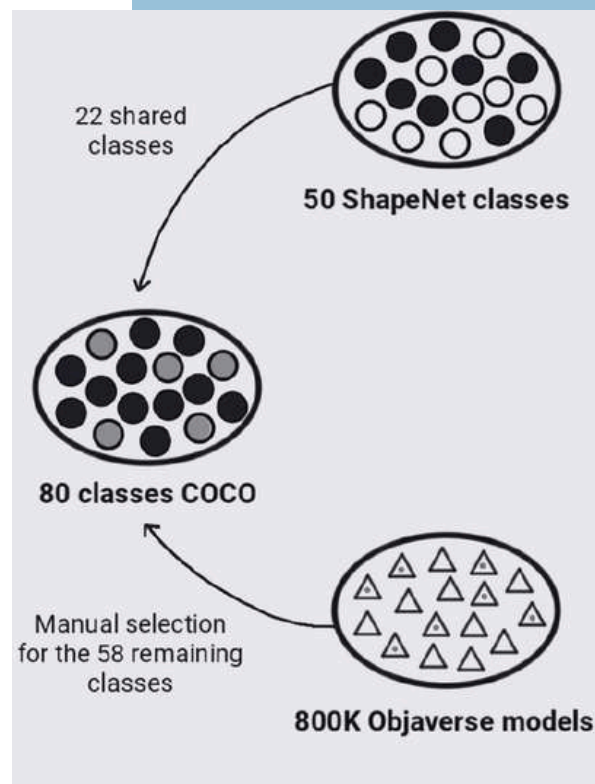
This summary is based on the paper 3D-COCO: Extension of MS-COCO Dataset for Scene Understanding and 3D Reconstruction *by Maxence Bideaux, Alice Phé, Mohamed Chaouch, Bertrand Luvison, and Quoc-Cuong Pham, presented at the IEEE International Conference on Image Processing (ICIP) on October 27th–30th, 2024, and available at* arXiv:2404.05641.

KEY FACTS

Coordinator: Engineering Ingegneria Informatica S.p.A. (Italy)
Start date: November 2022
Duration: 24 months
Consortium: 25 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

in @lago-europe   ✕ @lago_europe